

Assessment in Multilingual Schools

A comparative mixed method study of teachers' assessment beliefs
and practices among language learners - CLIL and migrant students

Helena Reierstam



Assessment in Multilingual Schools

A comparative mixed method study of teachers' assessment beliefs and practices among language learners - CLIL and migrant students

Helena Reierstam

Academic dissertation for the Degree of Doctor of Philosophy in Education at Stockholm University to be publicly defended on Friday 2 October 2020 at 10.00 online via Zoom, public link is available at the department web site.

Abstract

This thesis presents the results from two research projects on teachers' assessment beliefs and practices in multilingual education. Study I involved teachers of biology, history or English as a foreign language (EFL) in Swedish upper secondary *Content and Language Integrated Learning* (CLIL) schools, grades 10-12, where English is used as the medium of instruction. In Study II teachers from schools with newly-arrived migrant students (NAS) grades 7 to 12 in the natural and social sciences participated. In both contexts, students are learning the language of instruction at the same time as they are expected to develop subject area knowledge why issues in relation to the role of language in assessment come to the fore. The aim is to contribute to the knowledge of an underexplored research area on subject matter assessment in multilingual schools and draw attention to the consequences varying language policies and pedagogies may have on fairness in access opportunities and validity in assessment outcomes.

In this thesis teachers' language beliefs and practices as expressed in interviews, questionnaires and assessment samples were compared and analyzed in relation to the cognitive and linguistic requirements of language functions in syllabi and the assessment tasks. Whereas Study I was mainly qualitative in nature, involving 12 teachers, a mixed method approach was adopted in Study II where 196 teachers participated in a survey and 13 in follow-up interviews. The responses in the survey were analyzed using descriptive and inferential statistics. Thematic content analysis was used for the interview data, the assessment samples and open-ended survey questions.

The findings from the two studies indicate that although teachers state that language is not part of the assessment, they maintain that students need to use language to show proof of critical thinking and more advanced analytical skills. All teachers regardless of discipline shared the weight attached to covering course content as expressed in syllabi which points at a shared accountability culture. One of the main concerns expressed by the teachers is how to construct assignments where language does not represent a hindrance to show content knowledge. Teachers in both studies claimed to focus mainly on subject concepts, not general academic language, and the assessment beliefs and practices seemed to be closely related to the character of the subject. The non-parametric tests of association revealed that teachers with a dual language and subject content certification displayed significantly higher results in relation to all activities involving a visible language pedagogy, e.g. looking at useful sentence structures and providing model texts. Oral follow-up was used by all teachers to remedy poor written results. Although the use of the students' strongest language is advocated in guidelines for the instruction of NAS, most teachers in Study II referred to a monolingual Swedish language norm.

In a society where the educational discourse has become characterized by diversity, inequality and segregation, these two studies underline the need for a shared language policy and pedagogy across subjects and school contexts. They also suggest that an organization and teacher profession with an explicit responsibility for academic language is needed to provide equal access to subject content and validity and comparability in assessment in multilingual schools.

Keywords: *Assessment, content and language integrated learning, CLIL, newly-arrived students, teacher beliefs, language policy, academic language, subject content, fairness, equity, validity.*

Stockholm 2020
<http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-183910>

ISBN 978-91-7911-250-9
ISBN 978-91-7911-251-6



**Stockholm
University**

Department of Education

Stockholm University, 106 91 Stockholm

ASSESSMENT IN MULTILINGUAL SCHOOLS

Helena Reierstam

Assessment in Multilingual Schools

A comparative mixed method study of teachers' assessment beliefs and practices among language learners - CLIL and migrant students

Helena Reierstam

©Helena Reierstam, Stockholm University 2020

ISBN print 978-91-7911-250-9

ISBN PDF 978-91-7911-251-6

Cover illustration <https://www.shutterstock.com/sv/image-illustration/group-business-people-world-map-2523006>

Printed in Sweden by Universitetsservice US-AB, Stockholm 2020

Contents

Acknowledgements	x
Prologue – a licentiate and then a doctoral thesis	xii
Part 1 - Setting the scene	
Introduction	1
Background	1
Assessment, language and opportunities in life	1
Aims and research questions.....	4
An integrative comparative approach	7
Outline of thesis	10
Conceptual framework	12
Mono-, bi-, multi-, pluri- and translingualism	12
Language policy	14
The academic language(s)	16
Language and assessment in the subject disciplines.....	19
Assessment of language learners.....	22
Teacher beliefs and conceptions	28
Analytical models	35
Cummins’ matrix.....	36
The content-language integration continuum	38
The assessment continuum.....	39
The assessment literacy model.....	40
The language integration matrix.....	41
The curriculum alignment model.....	44
Assessment validation	45
Contextual background – The Swedish setting.....	46
Multilingualism	46
Education	48
Teachers and their professional development	51
Language in curricula and policy documents	52
Assessment and grading.....	55
Comparing CLIL and NAS	57
Summative comments on the contextual framework for teachers’ assessment in Swedish multilingual schools.....	63
Part 2 - Assessment in the CLIL context	
Study I: Assessing Language or Content?	69
Background	69
Aims and research questions.....	69

CLIL in Sweden	70
Previous research on CLIL and assessment	71
Methods and material	72
The semi-structured interviews and the questionnaire	72
The assessment samples	72
Participants and sampling	73
Findings in Study I	75
Language and content in the courses.	75
Assessment	77
Discussion	87
Concluding remarks	89
Part 3 - Assessment in the migrant context	
Study II: Assessment in Multilingual Schools	93
Background	93
Aims and research questions.....	96
Literature review and previous research	97
Methods and material	103
The ontological and epistemological approach of the study	103
Mixed methods design.....	105
Sampling process and participants.....	107
The instruments	116
Data collection procedure and data types	118
Analyses of data	121
Validity, reliability and trustworthiness.....	125
Ethical considerations	126
Findings in Study II	127
Language composition of the groups and teachers' experiences with NAS	130
Teachers' perceptions of the students	133
Teachers' experiences and beliefs about language in education	142
Teachers' assessment beliefs and practices across subjects and schools	150
Comparison across teacher groups	173
Professional development in assessment and language teaching methods	183
Summative comments about teachers' assessment among NAS	186
Discussion	189
Assessment beliefs and the purposes of assessment	190
Assessment and beliefs about students, language and local organization	192
Variation in teacher beliefs and practices	194
Concluding remarks - implications for policy and pedagogy.....	198
Limitations of the study	201

Part 4 - Comparing contexts

Synthesis of results..... 205

Comparing CLIL and NAS—language by design or by default?	206
Implications for assessment.....	210
Teachers' perceptions of the students	212
What teachers assess.....	213
How teachers assess.....	213
Implications for language policy.	217
Perceptions of language use and language pedagogy	217
Linguistic hierarchy in Sweden	227
Implications for organization and the teaching profession	229
Suggestions for future research.....	233

Summary in Swedish/Svensk sammanfattning 235

Inledning och bakgrund.....	235
Syfte och forskningsfrågor.....	237
Forskningsbakgrund och teoriram.....	238
Metod och material	240
Resultat	242
Diskussion och syntes	245
Implikationer	247

References 249

Appendices..... 273

Appendix A: Missivbrev på svenska	273
Appendix B: Letter in English	274
Appendix C: Enkät på svenska	275
Appendix D: Samtyckesblankett intervju på svenska	292
Appendix E: Letter of consent in English.....	294
Appendix F: Exempel e-mail till lärare inför intervju	296
Appendix G: Example of Email sent to teachers before the interview.....	296
Appendix H: Table distribution of teachers across school types and subjects	297
Appendix I: Tables with results from non-parametric tests for NS and SS teachers	298
Appendix J: Tables with results from non-parametric tests for SS and SSLang teachers.....	306
Appendix K: Tables with results from non-parametric tests across school types....	314
Appendix L: Bar charts: frequency use of different assessment formats	326

List of figures

Figure 1.1 Illustration of the key to unlock future opportunities for global citizens, i.e. language competence and grades.....	2
Figure 1.2 Research focus and participants in the study	4
Figure 1.3 Three interdependent perspectives of integration in relation to content and language	8
Figure 1.4 The language policy cube.....	15
Figure 1.5 What are the factors that influence teachers' knowledge and beliefs, and hence practice?..	30
Figure 1.6 Background factors influencing teacher beliefs	31
Figure 1.7 Adaptation of the model of constructive alignment	32
Figure 1.8 Cummins' matrix from 1981	37
Figure 1.9 Desired progression in assignments and student proficiency.....	38
Figure 1.10 The content-language integration continuum.....	39
Figure 1.11 Continuum of teachers' conceptions about assessment.....	39
Figure 1.12 The assessment literacy model.....	41
Figure 1.13 Positions of the language integration profiles.....	43
Figure 1.14 Proportion of newly-arrived students, including background unknown, in compulsory school for different municipal units during the schoolyears 2008/09 to 2016/17	49
Figure 1.15 Summary of content in Part 1 (curriculum & policy) and areas still to be investigated (teacher beliefs and assessment practices)	63
Figure 2.1 Outline of Study I.....	70
Figure 3.1 Interactive mixed methods' design.....	106
Figure 3.2 Factors influencing teachers' assessment beliefs and practices.....	117
Figure 3.3 Teachers' appreciation of the average number of languages in their classes	131
Figure 3.4 Teachers' experience of language learners prior to receiving NAS.....	133
Figure 3.5 Teachers' experience of using assessment data from the national diagnosis, "mapping".	134
Figure 3.6 Teachers' attitudes with regard to the importance of different assessment types for grading	156
Figure 3.7 The interviewed teachers' assessment profiles	191
Figure 4.1 Illustration of additive versus subtractive bilingualism	206
Figure 4.2 Illustration of additive bilingualism in different language registers (CLIL vs. NAS)	207
Figure 4.3 The interviewed teachers' language integration profiles	218
Figure 4.4 Alignment of languages in education with intended learning outcome/target language...	224
Figure 4.5 Language processing stages towards the target language of schooling	224
Figure 4.6 Comparative curriculum assessment in Swedish CLIL and NAS contexts	226
Figure 4.7 The Linguistic hierarchy of Sweden	227
Figure 4.8 The position of Swedish CLIL and NAS contexts along the Language integration continuum.....	230

List of tables

Table 1.1 Overview of research focus, time frame, participants, scope and methods.....	9
Table 1.2 Content-obligatory versus content-compatible language of schooling.....	17
Table 1.3 CDF (cognitive discourse functions) types and their members	20
Table 1.4 Analytical models used in relation to different parts of the study	35
Table 1.5 Features relating to the assessment profiles.....	40
Table 1.6 Four different approaches to integration of language.....	42
Table 1.7 Four possible language integration profiles.....	43
Table 1.8 Curriculum indicators	44
Table 1.9 Examples of CDFs and number of appearances in the knowledge requirements for E, the lowest passing grade in Swedish subject syllabi for upper secondary school, Grade 10	54
Table 1.10 Overview of differences and similarities in the Swedish CLIL and NAS contexts.....	62

Table 2.1 Participants in Study I.....	73
Table 2.2 Participating teachers' background	74
Table 2.3 Function words used in the test items.....	84
Table 2.4 Question words used in the test items.....	84
Table 2.5 What teachers claim to be most important when assessing students' skills	85
Table 3.1 Classification of Swedish municipalities (according to SALAR, 2017)	109
Table 3.2 Overview of response rate in the web-survey.....	110
Table 3.3 Stated cases of municipalities and number of respondents in the survey (exact numbers)	110
Table 3.4 Teacher background variables	113
Table 3.5 Main background features of the interviewed teachers	115
Table 3.6 Number of schools and municipalities contacted for the survey by month	119
Table 3.7 Duration of interviews with teachers in minutes	120
Table 3.8 Teachers' assessment samples	121
Table 3.9 Overview of handling of data and analytical procedures	124
Table 3.10 The subject disciplines of the interviewed teachers.....	129
Table 3.11 The school levels of the interviewed teachers	130
Table 3.12 Teachers' reported length of experience with NAS in their classes	132
Table 3.13 Perceived level of difficulty in different oral and written assignment formats	135
Table 3.14 Perceived level of difficulty in different language registers	136
Table 3.15 Perceived level of difficulty in different subject-specific skills (CDFs or function words)	137
Table 3.16 How teachers prepare the students for assessment tasks	143
Table 3.17 Who the teachers collaborate with in different language-related tasks	149
Table 3.18 What teachers believe is important for students to know for assessment in their subject	151
Table 3.19 Frequency of teachers' corrective feedback on students' language.....	152
Table 3.20 Teachers' perceived importance of different assessment types for grading	156
Table 3.21 Frequency in using different assessment types.....	158
Table 3.22 Local procedures with national tests.....	159
Table 3.23 Perceived language policy in assessment at the schools.....	160
Table 3.24 Teachers' language policy in assessment	162
Table 3.25 Accommodations provided by teachers' in the assessment tasks.....	163
Table 3.26 Comparing NS and SS teachers' use of assignments and perceived language difficulty ..	174
Table 3.27 Comparing NS and SS teachers' perception of difficulty in subject skills.....	175
Table 3.28 Comparing NS and SS teachers' perception of the importance of language features	175
Table 3.29 Comparing NS and SS teachers' use of different assessment formats and types	176
Table 3.30 Comparing NS and SS teachers' use of accommodations	177
Table 3.31 Comparing SS and SSLang teachers' use of assignments, perceived difficulty in assignments and language genre	178
Table 3.32 Comparing SS and SSLang teachers' use of assignments, accommodations and perceived importance of different language genres	179
Table 3.33 Comparing SS and SSLang teachers' preparation of students for assessment	180
Table 3.34 Teachers' self-reported professional development needs.....	184
Table 3.35 Teachers' comments on the major challenges of assessing NAS's content knowledge ...	187
Table 3.36 Alignment between intended learning outcomes, teaching and learning activities in class and assessment	199
Table 4.1 Teachers' self-reported assessment beliefs and practices in the two contexts (CLIL and NAS).....	214

Abbreviations

BICS – Basic Interpersonal Communicative Skills

CALP – Cognitive Academic Language Proficiency

CC – Content compatible (language)

CDF – Cognitive Discourse Function

CEFR – Common European Framework of Reference for Languages

CLIL – Content and Language Integrated Learning

CO – Content obligatory (language)

CR – Constructed-response (questions)

EFL – English as a Foreign Language

EMI – English-Medium Instruction

IFAU – the Institute for Evaluation of Labour market and Education Policy, *Institutet för arbetsmarknads- och utbildningspolitisk utvärdering* in Swedish

ILO – Intended Learning Outcomes

L1 – Language 1, i.e. the first language

L2 – Language 2, i.e. a second language

MC – Multiple-choice (questions)

OECD – the Organisation for Economic Co-operation and Development

NAE – the National Agency for Education, *Skolverket* in Swedish

NAS – Newly-arrived Students

NS – Natural Sciences

SALAR – Swedish Association of Local Authorities and Regions, *Sveriges Kommuner och Regioner (SKR)* in Swedish

SPSS – Statistical Package for the Social Sciences

SR – Selected-response (questions)

SS – Social Sciences

SSI – the Swedish School Inspectorate, *Skolinspektionen* in Swedish

SSL – Swedish as a Second Language

TA – Thematic (content) analysis

TLC – the Teaching and Learning Cycle

Transcription conventions

[...] = text has been deleted

[text] = added information, clarifications when a sentence is incomplete

... = pause or trailing off of speech

TEXT = louder voice used by speaker

Acknowledgements

One day it hit me, in the midst of conducting analyses, finding new references and writing up my results, this PhD process reminds me of “Labyrint”, a children’s program I watched on Swedish TV with my kids when they were younger: Three children competing against the figure Daidalos and his fantasy creatures in a labyrinth, earning life-pucks in various challenges. They had to cross a slimy puddle by setting foot on different slippery objects, struggling to find their balance, reaching the other side as best they could without falling to avoid an otherwise awaiting slime-shower. In my case there was no threat of “slime-shower”, but the labyrinth was there, the chasing creatures (read deadlines) were real as were the “slippery challenges”. Now the puddle has been crossed and I can note that it has been challenging but also extremely rewarding and fulfilling. The multidimensionality of this study has sometimes been overwhelming, but I would not have done it without the deviations and extra miles, I have learnt so much. Many times, I thought the search for more information or “in depth knowledge” would give me the “missing piece”, instead I ended up with an “additional piece” I didn’t even know existed. No matter how, I believe I earned a lot of “life-pucks” along the way. I would like to paraphrase and translate the words of a colleague (Maria Lim Falk) who described her doctoral trajectory very eloquently and turn them into mine: For a long time, I felt this is LIFE. Many a time over the last year I have noted with certain despair, this IS life. Now I know, this was PART of life and that is all it was.

Although the data-collection, analyses and writing were conducted as an individual task, many people have come alongside and contributed in so many ways. I would like to express my gratitude to my supervisors during the second study, Ulf Fredriksson, Eva Eliasson and Meeri Hellstén at the department of Education at Stockholm University. Ulf, my main supervisor, thanks for your solid and ready advice along the way. Thank you also for handfast support in maneuvering the final intense stretch without questioning the ravaging speed, at least not expressly. Our shared appreciation for “fika” does not have to end here. Meeri, I so appreciate you pushing me to network and offering me opportunities to connect with the world out there. Thank you to the two of you, Ulf and Meeri, for the great learning experience in Tokyo with colleagues at Tokyo university. Arigato daimas! Eva, I will always remember our first meeting, when you were waiting to welcome me in the hallway one of the first days, sharing your spontaneous excitement about my research topic. That positivity was good fuel all the way, thank you. Special credit to Staffan Nilsson at IPD, Stockholm University, for all knowledgeable help during the statistics consultations. Your help was invaluable and timely. Thank you Carina Carlhed Ydhag for your tireless and kind assistance with my questions and for offering initial support with SPSS, reference books and practical matters to the very end. Christina Edelbring, thank you for an open door, to sneak by and vent for a moment.

To my supervisors during FRAM-doctoral school and Study I at Gothenburg University, Liss Kerstin Sylvé and Gudrun Erickson. I thanked you back then when I finished my licentiate thesis, but I thank you again. You have been a constant source of inspiration in your respective fields of research and with your expertise. Thank you for your special support during that time, but also for continued inspiration and collaboration!

To my FRAM-colleagues, Linda Borger, Kent Fredholm, Maria Frisch, Tina Gunnarsson, Maria Håkansson Ramberg, Lisa Källermark Haya, Karina Pålsson Gröndal, Céline Rocher Hallin and Linda Smidfelt, thank you for staying in my life and being friends, colleagues and a continued sounding board. Our paths did not part, they continue to intertwine. Thank you to the CLISS-team in Gothenburg and elsewhere, Liss Kerstin Sylvé the research leader, Inger Lindberg, Ylva Sandberg, Per Holmberg, Tore Otterup, Eva Olsson, Elisabeth Olsson, Britt Marie Apelgren, Sölve Ohlander, BethAnne Paulsrud and Maria Lim Falk. Being part of your group is an honor and privilege. Special thank you to Liss Kerstin for bringing the CLISS anthology *Investigating Content and Language Integrated Learning. Insights from Swedish High Schools* into fruition. It is my first real book and is a major goal I fulfilled since claiming as a 7-year old that I wanted to be a writer.

My thanks also go to Britt Marie Apelgren and Max Scheija, my ninety percent readers, for your thorough reading, insightful suggestions and questions. Thank you to all my doctoral colleagues at IPD at Stockholm University 2017-2020, it has been a joy to know you, to come alongside and share in each other’s projects, ups and downs. I am especially grateful to Carles Fuster Sansalvador for good chats when we were sharing office and taking the

SPSS course in Gothenburg together. I cherish our joint interests and passion for languages as the “outliers” at the department of education. Reed Curtis and Brendan Munhall, your native language skills came in very handy during the final stage, thank you very much for your valuable input. Lovisa, Inga, Grant and colleagues at Stockholm University’s Avhandlingsstöd, thank you for providing solid and calm support in the final stage. Special tribute to my American friend, fellow “data nerd” and doctoral partner in crime, Kristine Thibault. You have saved me many times, sending me the SPSS handbook and meeting over WhatsApp, wrestling through this together on opposite sides of the “pond”. I know that there have been encounters and exchange with people I have not mentioned by name, but thank you all for providing important input, even in ways neither you nor I understood at the time. I dedicate this to all my friends as well, some of you of migrant background. We are in this world together.

Special thanks to my colleagues at the Unit for National tests at the National Agency for Education. I have been at the best place among the best colleagues during this journey. The lunch dates and fika chats have kept me grounded and contributed to my sanity. Thank you Karin Hector Stahre and Anders Boman for allowing me to work with the National Assessment framework. It has sharpened and developed my understanding in uncountable ways in collaboration with knowledgeable colleagues at the Agency and test developers at the Swedish Universities, to whom I want to extend a special thanks: You have enriched my knowledge base about assessment across disciplines IRL.

This thesis would not have been possible without the participation of the teachers in my study. Thank you for generously sharing your time, important experiences and perspectives. I also want to thank the NOSI-network, especially the facilitator and leader Carmen Galian Barrueco for inviting me and thereby enabling me to share your group’s valuable experiences and comments on my questionnaire. Maria Wiksten, thank you for taking time to review my questions. Many thanks go to those who provided me with extra funding, Knut and Alice Wallenbergs resebidrag and Rhodin’s donationsstipendium which allowed me to travel around the country for the data collection and to travel to Taiwan and the EASE (East-Asian Association for Science Education) conference along with other conferences.

Finally, my family: Adrian, maybe mom can finally work up her manual skills in “Human fall flat” without literally falling flat. I have felt like that limp avatar many times and it has given me perspective on the challenges involved in learning new skills as an adult. Thank you for providing moments of escape into your digital worlds of computer games and the books where we have explored multiple story treehouses, travelled to the little house on the prairie, to the banks of Plum creek and the shores of Silver lake. I look forward to many more books, and I guess games too, as long as you will let me. Alina, I remember your comment when I was coming out of the sauna, “Mom, did you come up with any new great ideas?” You had seen me coming out of there excited before, looking for pen and paper. Yes, the sauna has been a good incubator, and therapy for sore shoulders. Now that this dissertation is flying on its own, I am happy you are not, not quite yet. I long for many good moments when mom is no longer busy “writing her book”, as do you, I am sure. Thank you for all little notes of affection and affirmation on my table to spur me on, it has meant the world to me. Göran, you might be the only one in the family working the night shift from now on and you finally get rid of your wife’s innumerable piles and can vacuum without too much criss-crossing, although, I cannot make any promises. Love you my crazy energetic “cookie monster” and partner in crime. Thank you Dad, Ingemar Wickström, whose support was invaluable during my first study, and who passed on to better pastures before I finalized Study I. Mom, Birgit Wickström, you have left me gradually during this last study, but how I have cherished the moments of pause with you during this season, encouraged by your warm familiar smile. I recently realized how your constant encouragement to learn languages (Dad) and to care for students with learning disabilities (Mom) have led me to this day. I am so fortunate. To conclude:

“Not because of who I am, but because of what you’ve done,
Not because of what I’ve done, but because of who you are.”

Casting Crowns

Vallentuna, August 2020
Helena Reierstam

Prologue – a licentiate and then a doctoral thesis

My role as a researcher began in 2012 when I started my doctoral studies, leading up to a licentiate degree in 2015. After 14 years in Sweden and the US as a teacher of English and French as a foreign language and English for specific purposes, I was more intrigued than ever as to how to align curricular goals and my teaching with the most appropriate assessment methods to best serve and motivate the students. I was struggling particularly with the *how* and *what* of assessment. Language learning theory and course syllabi were characterized by the communicative language approach, as in producing meaningful and real communication. The focus was function and use, more than form. Which assessment types would be most authentic? And what kind of language domains should be in focus? Back in the days when I earned my teaching degree, assessment was not part of the teacher education curriculum. Language teachers were supposed to use photocopiable tests found in teachers' manuals from the educational publishing companies and national tests. Other than that, teacher colleagues seemed to repeat the assessment practices they had been exposed to themselves as students, developing their own tests such as vocabulary quizzes, grammar tests, translation assignments and essays, which were not necessarily very communicative or authentic.

One day I accidentally overheard a conversation in a grocery store where two teachers were talking about how their entire spring semester meant what could be labelled as “teaching to the test”; in other words, practicing old national tests before administering and grading that year's national tests. I recognized this as something my colleagues and I did too, but it struck me as odd when I heard these teachers speak about teaching their students “test-wiseness.” My teaching experience after a couple of years as a foreign language teacher in an American high school north of Boston added to the confusion, prompted as I was to construct multiple-choice items for the finals in French. This led to my initial research proposal in 2012, aimed at mapping language teachers' assessment practices in Sweden. However, an attentive supervisor noticed my cross-disciplinary interest in integrating subject content into the language courses (sometimes referred to as content-based language teaching, or CBLT) and subsequently I was invited to join an ongoing research project at Gothenburg University on content and language integrated learning in Swedish schools (CLISS). I contributed an investigation on assessment in content and language integrated learning (CLIL). It was the ideal match for me.

When I started my research agenda in 2012, I was inexperienced as a researcher and had much to learn about the multifaceted field of assessment in different subjects and language across the disciplines. An interdisciplinary study covering three subject disciplines—biology, history and English as a foreign language—with very different assessment traditions and epistemological perspectives challenged me. Theoretically I had little insight into the didactic foundations of biology and history. Methodologically I invented the wheel along the way, and I had not yet pondered paradigm wars and dilemmas in mixing methods. As I re-embarked in 2017, the preparedness was somewhat higher. I was privileged to have worked cross-disciplinarily with validation plans and assessment frameworks for national tests in all subjects together with knowledgeable colleagues at the National Agency for Education and experienced test developers at different Swedish universities. The initial fascination and mixed emotions in relation to assessment had thus brought me from teachers' summative assessment practices in foreign languages, via assessment and language policy in CLIL, to issues of equity, fairness and teachers' language awareness in subject assessment in multilingual educational settings. How and why I choose to use the term “multilingual” will be outlined in the following.

The first year back as a doctoral student opened my eyes to more of the theoretically diverse landscape and the positioning of different stances, which will become apparent in the following. However, the most pressing motive to go back into research was my impression of an overarching need to bring attention to what consequences linguistically-diverse schools, where every teacher is now in a sense both a language teacher and language assessor, have on assessment, the status of languages and the intended learning outcome as expressed in policy documents. The work at the National Agency made me aware of differing conditions in relation to language policy and the status of languages in instruction for newly-arrived migrant students and English-medium instruction. In the midst of good intents, I noted a certain confusion and incoherence regarding motives and suggested steps to be taken in relation to language-learning goals in curriculum and policy documents. In certain educational contexts where English is used as a medium of instruction, no explicit policy concerning language was expressed, whereas recommendations to use translanguaging were forwarded in other linguistically-diverse settings among migrants. All of this took place without much talk about the ideology behind different approaches to language learning and assessment, or alignment between purpose, pedagogy and its consequences. I sensed conflicting agendas between how language was conceptualized and how assessment was perceived.

As a former teacher with experience of both homogenous and heterogeneous student groups, I see students as loaded with capabilities and potential, and teachers as facilitators and expert guides. Benjamin Zander, a conductor of the Boston Philharmonic and a teacher, as well as the author of a 2002 American bestseller, said in his book that his job is to “awaken possibility in other people,” which I think is beautifully put. He also said: “In the measurement world, we set a goal and strive to achieve it. In the universe of possibility, we set the context and let life unfold.” The quote suggests a dichotomization which can be found in certain paradigm frictions, where measurement, and thereby assessment, is considered evil. With that I disagree, as long as the goals are realistic and achievable. I would also find it disrespectful to students to say “we set the context,” as suggested in the quote above. Unfortunately, there are circumstances which cannot be changed, or else why would people feel urged to leave their homes and put themselves in the predicament of “minority groups?” My hope for all students is that a wonderful life of possibilities will unfold, but I think it helps to set goals to get there. However, that requires providing students with the best equipped professional guides, which I believe can change their context.

All of this is to say that what started as an unsettling feeling never ended. My curiosity about the status of different languages in education, the consequences this may have for students in different contexts, for instruction, in relation to the intended learning outcomes and fairness in assessment was never satisfied. Moreover, assessment appeared to be a blind spot where more research was needed. So here, off I go, hoping to contribute and shed some light on the diversified landscape and help gain insights that may somehow help build bridges to more equitable assessment policies and practices in a linguistically diverse and global school.

Part 1 - Setting the scene

Introduction

Fair assessment is inseparable from fairness in access opportunities and in what the curriculum offers. (Stobart, 2005, p. 278)

Background

Assessment, language and opportunities in life

Equity, equality and fairness are key features of quality education and cannot be considered in isolation from curriculum, assessment and educational opportunities for the students. However, according to recent reports from the Organisation for Economic Co-operation and Development (OECD, 2019a), inequalities across schools and students are increasing in Sweden and the gap between advantaged and disadvantaged students is wider than the OECD average. While more compulsory school students attain the highest grades, more students are also found at the other end of the spectrum (SALAR, 2016). In spite of cultural diversity in society at large, this is not equally reflected in the schools and classrooms. Instead, children with similar backgrounds are increasingly segregated into the same schools. Already in 2017, the OECD recommended measures to promote integration, raise outcomes for immigrant students, improve working conditions for teachers to motivate the most competent to stay in the profession, and ensure comparability of assessment results across schools. In 2019, suggestions along the same lines were made, to change the enrollment to private schools to avoid segregation, but also to improve teacher education and raise the status of the profession by strengthening continuous learning for teachers and instigating more collaboration between colleagues (OECD, 2019a).

Inequality in grading between schools and teachers has been recognized as a problem and an area needing improvement in several national investigations. The Swedish School Inspectorate states in a report (2018) that teachers in many schools lack important conditions to ensure fair and equitable assessment. There is also a lack of time to calibrate assessment with colleagues, at the same time as there is a shortage of certified teachers and also competition between schools that can be traced in discrepancies between results on national standardized tests and more favorable final grades. The School Inspectorate notes that it is crucial that the students' knowledge is assessed on equal grounds, since teachers' decisions when grading may have consequences for the students' opportunities for continued studies and professional goals. From an international perspective, Menken (2008) critically addresses the enormous power tests and assessments wield over the lives of students, especially minority groups, who are the most vulnerable to inequitable educational decision-making based on test scores. Since major decisions are made in schools every day and intersecting research between testing and language policy is fairly recent and scarce, more research that can offer guidance for the use of assessments is needed (Menken, 2008, p. 410). In a meritocratic society, assessment of educational achievement and knowledge becomes the key to equal opportunities (Gipps & Murphy, 1994); however, equal opportunities can only be achieved if the assessment is valid and fair. As noted by Young (2009), the primary concern in evaluating validity is fairness, and the measure of fairness is comparability.

According to a research overview from the Swedish Research Council (Forsberg & Lindberg, 2010), relatively few studies have examined teachers' assessment and grading practices in their subjects.

Applied research integrating knowledge theories and assessment, and investigating how research, policy and pedagogy interact in a local Swedish context is needed, according to the authors. However, some studies have been conducted subsequently (cf. p. 56f), but research in relation to assessment in subject matter disciplines is still limited. In Sweden, national standardized tests are used in many subjects to give students the same opportunity to show their knowledge. The purpose of the tests is to support teachers in providing just and fair assessments and grades, according to the website belonging to the National Agency for Education (henceforth called the NAE; the name of the agency is Skolverket in Swedish). Nevertheless, a student's grade (and thus opportunities) does not rest solely on the results of a national test if there is a national test in the actual discipline that year (see p. 55). Students' grades also rest on individual teachers' assessment practices and their interpretation of national curricular goals. Cultural and linguistic diversity render assessment more complex, especially in relation to issues of fairness and access. In an NAE policy document called Education for Newly-Arrived Students (2016), it is stated that teachers may need further training in assessment and language development to promote the highest possible academic success of newly-arrived immigrant students. Assessing language learners in a multilingual setting requires attention to certain challenges. Fairness in education means that all students must be offered equal access to the curriculum and examinations and assessment must be made as fair as possible for all groups. Assessment that overlooks issues of diversity and equity contributes to inequalities in outcomes (Bal & Trainor, 2016), whereas "if assessment is reliable, valid and fair for all students from start to finish, then it can serve as the bridge to educational equity" (Gottlieb, 2016, p. 1).

In an increasingly global world, international mobility and migration have clear implications for the school-age population and thus for education, learning outcomes and assessment. However, even if linguistic and cultural diversity may be perceived as problematic in relation to assessment, seen from a different perspective, multilingualism is desirable. As it says on the European Commission's website when promoting language learning: "languages unite people, render other countries and their cultures accessible and strengthen intercultural understanding" (Bonjean, 2018). In the same vein, it is noted that foreign language skills play a vital role in enhancing employability and mobility, as well as improving the competitiveness of the economy.

If language is the key to access in a multilingual society, grades are key to opportunities in life. Without language and valid learning outcomes some doors may never open. The metaphor of keys to different opportunities in life is used in the illustration below.

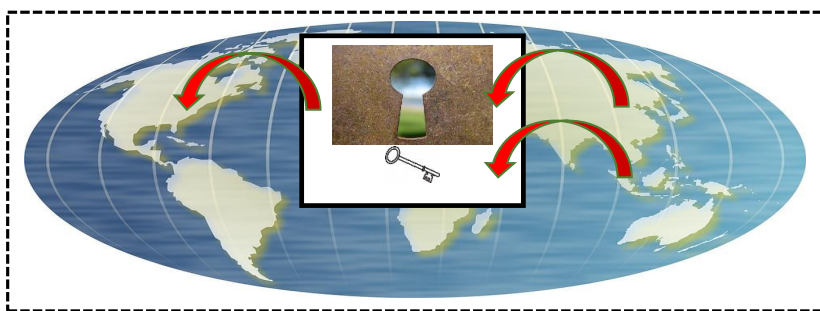


Figure 1.1 Illustration of the key to unlock future opportunities for global citizens, i.e. language competence and grades

The big frame represents the world and the smaller box in the middle represents a narrower national context like Sweden, for instance. While migrants from the outside world are arriving, seeking keys for inclusion and opportunities in Sweden, Swedish students are looking for keys for departure and thereby increasing their future options. In an educational context, teachers are the key holders. Part of their job is to help equip students with the necessary skills and knowledge, while also taking on the role of judge and gatekeeper. As assessors, they are responsible for making decisions to support teaching but also for generating data for public reporting and selection purposes, decisions for which they can be held accountable. The illustration comprises a linguistically-diverse educational landscape (seen through the keyhole) in which there is a main actor, the teacher. The professional role of the teacher requires certain assessment competencies in order to accurately evaluate student performance and make accommodations when necessary for better student attainment. The professional skills to plan, implement and interpret assessment data are sometimes referred to as assessment literacy (see p. 33). As stated in the government's official investigation (SOU 2017:54, p. 21), improving newly-arrived students' eligibility for upper secondary education requires a lot of the teachers, both in terms of knowledge and experience.

Against this backdrop, the motivation for the present thesis relates to fairness, equity and validity in education. Fairness is a complex issue with alternative definitions depending on emphases and stance; fairness can either be regarded as a component of validity or vice versa (Kane, 2010). In this study I adhere to the view expressed by Kane (2010, p. 177f): they are intertwined, since the overlap between fairness and validity is more pronounced than the differences. They are hence seen as closely related ways of looking at the same basic question, if the proposed interpretations and uses of tests scores are appropriate for populations from different contexts. Kane defines two types of fairness: *substantive fairness* and *procedural fairness*. The first implies that test takers should have equitable opportunities to learn and thus be able to earn the same average test score with respect to the construct a test is intended to measure and the second, that everyone should be treated in the same way, which is the case with standardized tests for instance. The question is fairness for whom? Using the definition of Stobart (2005), equity represents the judgement about whether equality in the form of opportunity and/or of outcomes has achieved just and fair results.

A founding principle in the Swedish education system is "the compensatory mission" (A. Andersson et al., 2015; Nilsson Folke, 2017), which aims at providing necessary support for equal opportunities to succeed. This becomes especially challenging in a heterogeneous, multilingual setting among newly-arrived students or language learners in general. Since educational assessment hinges on decisions made by authority practitioners, in this case the teachers, it is vital that their decisions are based on well-founded, conscious, consistent and valid assessment practices. People who use assessments need to be able to demonstrate that the assessments are justified and be held accountable to stakeholders (Bachman & Palmer, 2010).

The present study investigates teachers' self-reported assessment beliefs and assessment practices in two bi-/multilingual Swedish educational contexts among second-language learners (the distinction between bilingual and multilingual will be further outlined below). As an introductory declaration regarding the use of the term multilingualism in this study, I refer to a definition in which it is conceived of as a generic term including the use of two or more languages and in which bilingualism can be included (Aronin & Singleton, 2008; Cenoz, 2013). The common denominator in the two contexts is the use of a second, or foreign language (for a definition, see p.4) when assessing the students' content knowledge, but the students' backgrounds vary considerably due to their educational backgrounds and previous schooling. The teaching contexts vary also between on the one hand, CLIL, where English is

used as a medium of instruction and on the other, education for newly-arrived students (henceforth called NAS) in Swedish schools. According to the Swedish Education Act (SFS 2010:800, Ch.3, §12a), a “newly-arrived student” is a student who was born in another country, started school after he/she turned seven, and has been enrolled in Swedish school for a maximum of four years. Even if there are differences between the two student categories and teaching contexts (a comparative table is provided at the end of Part 1), the assessment situation for the teachers includes several similar features in relation to access opportunities and validity in assessment, which is the main concern throughout this thesis. Echoing the words of Abedi (2008, p. 337), “As the numbers of language learners increase assessment equity and validity are becoming priorities for educational policy makers”; and, I might add, for educational practitioners as teachers, especially when grading is involved.

Aims and research questions

The aim of the present study is to contribute to the knowledge of the underexplored area of assessment in multilingual content and language integrated education by investigating teachers’ assessment beliefs and assessment practices regarding language learners. The intention is to describe and examine the policies and practices of language use in relation to the *what* and *how* of assessment and hence their possible consequences for fairness in access opportunities and assessment outcomes.

The study consists of two studies in one; or, put differently, one study in two parts. The first study was conducted from 2012 to 2015 and the second from 2017 to 2020. The overarching research focus is teachers’ assessment beliefs and assessment practices in a bilingual or multilingual context where students are asked to demonstrate content knowledge in a second language (henceforth called L2, as opposed to the first language learned, or L1). In this study, a second language is a language that is learned in a community where the language is commonly spoken by the majority of people; second and foreign languages are often labeled L2 (Oxford, 2017). A foreign language is a language learned outside the countries where the language is spoken (Richards & Schmidt, 2002). Figure 1.2 below offers an illustration of the research focus and the participants in the two studies.

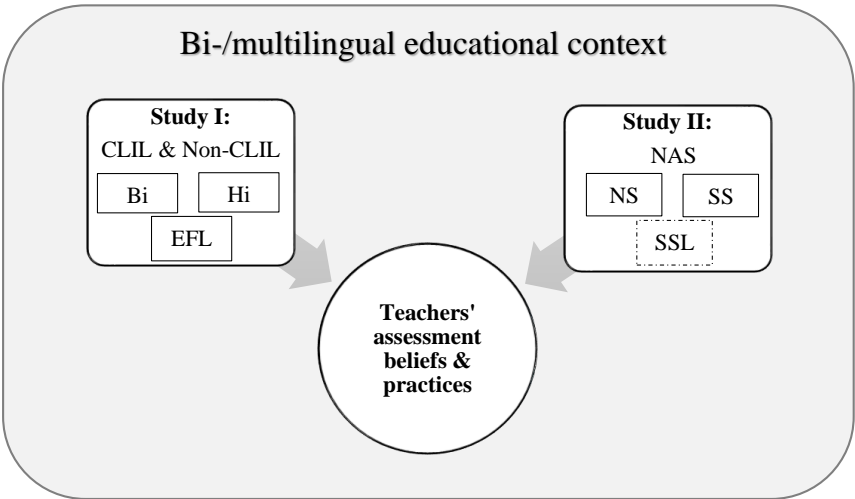


Figure 1.2 Research focus and participants in the study

The first bilingual context represents Swedish upper secondary schools with an intentional international CLIL profile, using English as the medium of instruction, where the students' first language (L1) is often Swedish. CLIL is advocated by the European Commission as a teaching approach to promote the learning of more foreign languages (Eurydice, 2006). Bilingual is used here since the term bilingual is often used in CLIL (Nikula, 2017; Pérez-Cañado, 2012), referring to a desire to build simultaneous skills in both the target and the native language. However, as noted previously, I use the term multilingual as a generic term, including the use of two or more languages (cf. Cenoz, 2013). The terms bilingual and multilingual are further outlined on page 12f.

The second multilingual context represents lower and upper secondary schools (Grades 7–12), where Swedish is used as the medium of instruction among newly-arrived immigrant students (NAS). Some of these schools have very little prior experience of linguistic diversity but, more importantly, bilingual instruction has not been a choice by the schools nor the students. The second study thus represents a linguistically more heterogeneous setting, as many different L1s may coexist in the same classroom. As previously noted newly-arrived students are defined in the Swedish Education Act, Amendment 2015:246 (SFS 2010:800), as students aged 7–18 who have migrated to Sweden and are considered to be newly-arrived for up to four years after starting Swedish school.

In Study I (see left box in Figure 1.2), teachers in two subject disciplines—biology (Bi) and history (Hi)—participated together with teachers of English as a foreign language (EFL). In Study II (right box), teachers in several subjects in the natural sciences (NS) and the social sciences (SS) were invited to participate (for a description of the sampling process, see methods and material, Part 3). Some of them also taught Swedish as a second language (SSL). For the sake of the first study, a distinction is made between CLIL and non-CLIL teachers due to the comparative approach concerning the potential effect the use of English as a medium of instruction might have on the CLIL teachers' view of assessment. Mainstream colleagues at the schools using Swedish as a medium of instruction are therefore labelled non-CLIL. The context among migrant students is generally not defined as CLIL (Cenoz, 2015), although common features have been acknowledged and will be discussed at the end of Part 1 and in the synthesis in Part 4. In Study II, the participants are teachers of newly-arrived migrant students (NAS). The purpose of the second study is not to compare teachers of NAS with mainstream colleagues, as in the first study among CLIL; however, after collecting data it became apparent that some of the participants made a distinction between their teaching of NAS in the introductory program or in preparatory classes versus teaching regular classes with only a few NAS.

A delimitation was made for the second study to focus on subject content teachers and not language teachers, as in Study I, where EFL teachers were included. However, although content teachers in the current disciplines were pinpointed, teachers were invited to note if they also taught Swedish, SSL, English or another language. The results show that some of the participants and teachers (in the social sciences) also teach SSL, and therefore this is included in the figure above. This is explained further in the section on methods and material. Another delimitation was to focus mostly on what can be described as summative written assessment, since this has implications for fairness in grading, although I consider formative and summative assessment to be closely interconnected (see p.21).

The two studies were conducted in Swedish public schools. Although the two investigations were conducted five years apart, the assessment system was the same and had not changed during the time of the study. However, in 2011, a new revised national curriculum and new subject syllabi were implemented, as well as a new grading scale; this means that teachers in the first study were still processing the new policy documents. Five years later, teachers were more accustomed to the new developmental perspective on learning as reflected by the assessment system, in which learning is

conceptualized as progress toward higher levels of competence in a criterion-referenced system, (see p.55 on assessment and grading).

Below, the research questions for the two studies are presented. They will be discussed in more detail in Part 2 (Study I) and Part 3 (Study II). Again, the research focus is subject content teachers' beliefs about assessment and how the bi-/multilingual teaching context (CLIL vs NAS) affects how language is dealt with in assessment. The overarching questions refer to both differences and similarities, which are two sides of the same coin. As regards the two contexts (CLIL and NAS), differences and similarities have been posited as will be discussed in the following. However, embedded in the questions is a notion that assessment practices (overarching research question 1 below) may be guided by teaching contexts, whereas conceptions of language are guided by subject discipline and policy (question 2). The findings will show what kind of relationships hold.

Overarching research questions, Studies I and II:

- What are teachers' assessment beliefs and assessment practices regarding language learners? Do they differ depending on context, CLIL vs. NAS? If so, *how* do they differ?
- What are teachers' conceptions of language in relation to the subject and policy? Are there common features between CLIL and NAS? If so, *what* are they?

Research questions, Study I:

- CLIL vs. non-CLIL: do the assessment practices differ in the two subject content courses history and biology due to the language of instruction? If they do, *how* do they differ and *on what grounds*?
- Are the assessment tools and the course content affected in the English language courses where English is used in subject content courses? If so, *how* are they affected?
- What does the assessment design look like in the different disciplines when it comes to *language, content* and *form*? Are there common features?

Research questions, Study II:

- What are the subject content teachers' self-reported assessment beliefs and assessment practices regarding newly-arrived students?
- How are the teachers' self-reported assessment practices associated with their beliefs about the students and language policy in education (i.e., curriculum, syllabi, translanguaging practices and local organization)?
- Do teachers' assessment beliefs and assessment practices vary depending on school level and subject discipline? If so, *how* do they differ and *on what grounds*?

The research questions imply a mix of teachers' self-reported beliefs and practices as expressed in questionnaires, semi-structured interviews, and observations of teacher-developed assessment samples. Teachers' tacit knowledge is like a black box but can be made visible through self-reflection and self-

developed assessment tasks or tests. Conceptions of assessment and actual assessment practices have been found to be significantly associated (Brown, 2008).

An integrative comparative approach

This study is integrative, as it is fusing language in education with subject matter content. However, it is also integrative in its theoretical stance, combining perspectives from teacher beliefs and validity theory in the field of educational assessment. Moreover, it is interdisciplinary in combining pedagogical perspectives from the natural and the social sciences, as well as language education. Subject integration in education is multidimensional, as it brings teachers together with differing disciplinary traditions, epistemologies and beliefs. Whereas teachers' practices are affected by educational policy and teachers' disciplinary identities, they may also be affected by the character of the students and the teachers' personal backgrounds and experiences. Comparing and combining various subject disciplines thus means theoretical integration, combining more or less autonomous perspectives fueled by shared cross-disciplinary concerns for students' learning in multilingual educational contexts.

CLIL explicitly unites two entities: content and language. This suggests that the two are separate, whereas indeed they are inherently interdependent (Nikula et al., 2016, p. 2). Nikula et al. (2016) maintain that the composed nature and multidimensionality in CLIL practices (including institutional, educational, personal and pedagogical purposes) needs a correspondingly multidimensional conceptualization of integration. They suggest a threefold model comprising *curriculum and pedagogy planning*, *participant perspectives* and *classroom practice*, which can be translated into the *what*, the *who* and the *how* of education.

The model fits as a description of the components of the current study, where the first, curriculum and pedagogy planning, is substituted by curriculum and policy. Curriculum can be defined as course content and the comprehensive collection of learning experiences, sometimes defined as the implemented curriculum (cf. p.44) while at the same time it can represent the formal written standards that are used to ensure nationwide uniformity, in this study labelled as national curriculum. Policy is defined as "a set of guidelines that determine how one should proceed given a particular set of circumstances" (Bell & Stevenson, 2006, p. 14). Bell and Stevenson (2006) express that policy consists of organizational principles and operational practices that are informed by values. In educational policy, the guidelines can be expressed in national curriculum and subject syllabi containing educational standards and core objectives.

The second component, participant perspectives, has been specified as teacher perspectives for this study, more specifically as teacher beliefs (see p. 28). In the model below it is labelled as teacher. The third component, classroom practice, has been transformed into teachers' self-reported assessment practices (see Figure 1.3). Assessment is often perceived as an integrated part of teaching and classroom practices, a perspective I share.

The common denominator affecting all three components is language/content, whether seen as two discrete entities or as interdependent, thereby enclosing the entire model.

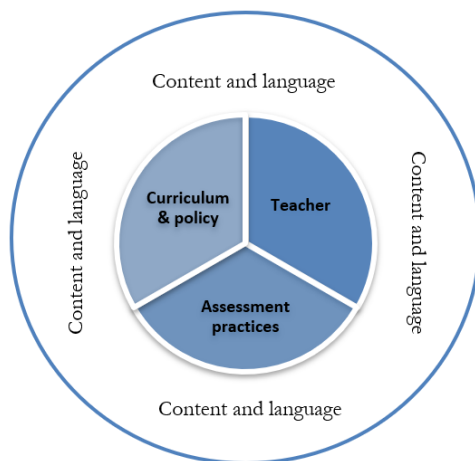


Figure 1.3 *Three interdependent perspectives of integration in relation to content and language*
Adapted from Nikula, Dalton-Puffer, Llinares & Lorenzo (2016)

If there is an order between the three, curriculum and policy would be placed first since intended learning outcomes take their start in curricular goals, as in constructive alignment (cf. p. 32). In curriculum studies, the concepts *intended curriculum*, *implemented curriculum* and *assessed curriculum* are often used (Porter, 2002). These curriculum levels are presented below in the Analytical models' section (p. 44) and the concepts are used in the synthesis of results. The key issue that is addressed in the current study is the under-researched area of assessment of second language learners' subject matter knowledge in a non-native language with monolingual national curriculum, and sometimes translangual policy.

The study is comparative as it is interested in finding out about differences and similarities by contrasting different teacher perspectives: e.g., CLIL with non-CLIL teachers, CLIL with EFL teachers, CLIL with teachers of NAS, social science teachers with natural science teachers, and social science teachers with or without a dual-language teaching certificate, to mention some of the most prominent. Comparative education is characterized as an interdisciplinary and multidisciplinary science, one that is approached and constructed from a variety of perspectives (Kazamias, 2009). Comparative research can be descriptive and analytical (as in this study), exploring relationships between variables while trying to understand why and how these can be interpreted, and going through the four stages of description, interpretation, juxtaposition before arriving at comparison. (Hamilton & Ravenscroft, 2018). Making comparisons to understand what works in education and what does not requires valid units of comparison; in this study, teachers' self-reported assessment beliefs are investigated in relation to policy and curriculum (Bray et al., 2007). Whether comparative research is concerned with issues at a global macro-level or national micro-level, Suter et al. (2019) conclude that what matters is how national and local leaders use the results of comparative studies to justify revisions of their educational policies.

In Table 1.1 below, an overview is provided of Studies I and II, including the shared overarching research focus, the data collection time frame and setting, the number of participants and subject disciplines, and the methods and materials.

Table 1.1 Overview of research focus, time frame, participants, scope and methods

Overarching research focus:	<p><i>Subject content teachers' beliefs about assessment and the role of language in bi-/multilingual teaching contexts (CLIL vs Newly arrived migrant students/NAS)</i></p> <p><i>Interdisciplinary influences/collaboration between subject disciplines and target language courses (English or Swedish)</i></p>			
Data collection: Timeframe and school form	<p>Study I (2013)</p> <p>Upper secondary, Grades 10-12</p>		<p>Study II (2018-2019)</p> <p>Lower secondary, upper secondary, introductory program, Grades 7-12</p>	
Participants: Teachers of...	<p>CLIL biology</p> <p>CLIL history</p> <p>English/EFL</p>	<p>Non-CLIL biology</p> <p>Non-CLIL history</p>	<p>Biology</p> <p>Physics</p> <p>Chemistry</p> <p>Swedish/SSL</p>	<p>History</p> <p>Geography</p> <p>Religion</p> <p>Social studies</p>
Methods: Instruments sequencing, and number of participants	Semi-structured interviews N=12		Questionnaire N=196	
	Questionnaire N=9		Semi-structured interviews N=13	
	<p>Written assessment samples * (N=42):</p> <p>Biology CLIL: n=4</p> <p>Biology non-CLIL: n=4 (+3)</p> <p>History: n=5 (+3)</p> <p>History non-CLIL: n=4</p> <p>EFL: n=14 (+5)</p>		<p>Written assessment samples * (N=26):</p> <p>Biology: n=1</p> <p>Physics: n=5</p> <p>Geography: n=5 (+1)</p> <p>History: n=1 (+1)</p> <p>Religion: n=5 (+2)</p> <p>Social studies: n=9 (+8)</p>	

*The parenthesis indicates other assignments used for assessment, e.g. prompts for group assignments, oral discussions or projects.

For more detailed information, see the methods and material sections in Part 2 and 3 where the data collection procedure, the sampling process, the instruments, the methods of analyses and participant background are presented.

Outline of thesis

The present thesis consists of four parts preceded by a prologue. The layout of the thesis is due to the integration of two studies: the first resulted in a licentiate thesis in 2015 and was followed by a doctoral thesis in 2020. The two studies—Studies I and II—outlined in Parts 2 and 3, respectively, are framed by an introduction in Part 1 providing the conceptual and contextual background of the two studies and Part 4, which presents a concluding synthesis of results. Below, a brief outline of the different parts is offered.

Prologue: The introductory prologue serves as a backdrop as regards my background and preconceptions as a researcher and shares the history and reasons behind the two separate yet integrated studies.

Part 1: Following an introduction regarding the importance of equitable assessment in linguistically-diverse educational settings, the aims and research questions are presented followed by a description of the theoretically-integrated nature of the study as well as the comparative approach. The conceptual framework section presents the theoretical and conceptual underpinnings of the main components of the study: i.e., language, assessment and teacher beliefs. First, some perspectives on language learning and language policy in education are presented. Included is a discussion of the ideological and ontological perspectives behind various notions of multilingualism. Next, language and assessment are delved into with definitions of academic language and a look at how content and language are integrated in course syllabi. A brief section follows on how language is dealt with in different assessment formats, leading to a section on assessment among language learners and possible accommodations. The teacher is in focus next, providing definitions and a discussion in relation to teacher beliefs and conceptions. This part moves from the teacher to how the teachers' backgrounds and knowledge have implications for the planning and alignment of instruction and, hence, assessment. I close with some comments on the importance of assessment-literate teachers. An overview of the analytical models used in both studies and the synthesis of results, which have been partly mentioned previously, closes the section.

Contextual background – The Swedish setting comes next, providing an outline of the Swedish educational context. A brief outline of the multilingual situation, a description of different school levels and types, and conditions for teachers' professional development introduce this section. Perspectives on language in education and policy documents are discussed, as well as the Swedish grading system, before concluding with a systematic comparison of features in relation to CLIL and NAS in Sweden.

Part 2: This part provides a summary and abridged version of Study I on teachers' assessment beliefs and practices in CLIL (Reierstam, 2015). The study on assessment in the Swedish CLIL context is one of the first on assessment in CLIL in Sweden and was part of a 3-year longitudinal project in Sweden called the CLISS project, or Content and Language Integration in Swedish Schools. The current summary implies minor developments, found in the inclusion of a couple of new references that appeared after 2015, together with some new quantitative comments. The full text is available in open access as *Assessing Language or Content? A comparative study of the assessment practices in three Swedish upper secondary schools*. A summary can also be found in the anthology following the CLISS project (Reierstam & Sylvé, 2019). For this study, a multi-method approach was used in gathering descriptive data for the qualitative analyses.

Part 3: This part covers Study II in the NAS context. First, a background is provided as to the current situation after the surge in NAS after 2013. A brief summary of previous research foci in the field is provided, together with a brief literature review. In the methodology section, the use of an integrative mixed-methods approach is outlined and discussed. Descriptive and statistical data are used in combination. The findings are first presented thematically, integrating and triangulating data from the

statistical analyses, the semi-structured interviews and the document analysis of assessment samples. At the end, a comparison is made in relation to different teacher groups and teachers at different school levels, outlining significant results from the non-parametric analyses. For the statistical analyses, the SPSS software, Statistical Package for the Social Sciences, was used. The section ends with a discussion in relation to the three research questions, suggestions for the future and some remarks regarding the limitations of the study.

Part 4: In the synthesis of results, the two contexts—CLIL and NAS—are described, juxtaposed and analyzed in integration with reference to the overarching research questions. Implications of the findings for language policy, assessment practices, local organization and the teaching profession are discussed. Issues of validity, fairness and equity that were raised in the introduction in Part 1 are explored and evaluated. Future research is suggested, and the thesis ends with a summary in Swedish, a list of references and appendices.

Conceptual framework

Following the three-dimensional integration figure presented previously (Figure 1.3), the conceptualizations of the study's intersecting fields (i.e., language in curriculum and policy, assessment and teacher beliefs) are outlined below. Starting with language, a brief background is presented as to the meaning and ontological implications of mono-, bi-, multi-, pluri- and translanguaging, together with an overview of language(s) in education. Thereafter, an outline of assessment in relation to the following is provided: assessment in the subjects; assessment of second-language learners (including test accommodations and aspects of validity). Then, a discussion on teacher beliefs and conceptions as used in this study follows, synthesized with teachers' assessment literacy.

Mono-, bi-, multi-, pluri- and translanguaging

"The limits of my language mean the limits of my world," says a famous quote by Ludwig Wittgenstein (1921). Living in a global world may instill a belief that only the sky is the limit. However, the fact that separate languages exist entails limitations, which become obvious when policy documents define when, how and what languages to use in various segments of society. The European Council notes that languages are not only a fundamental aspect of people's lives, but are also fundamental to the democratic functioning of society (Council of Europe, n.d.). One of the policy goals set forth by the European Commission after the Barcelona summit in 2002 presents a one plus two language goal, advocating that all citizens should learn at least two foreign languages besides the first language (L1) (European Commission, 2003). As defined previously, a foreign language is a language taught in school to be able to communicate with speakers of that language (Richards & Schmidt, 2002), whereas a second language is learned in a context where the language is spoken and is needed to be able to participate in society (Oxford, 2017). In regular foreign language teaching as well as in CLIL, where a foreign language is used as a medium of instruction in subject matter courses, the goal is additive bilingualism, where a second language is acquired without detriment to the first (Matthews, 2007), as opposed to subtractive bilingualism, where the new language replaces the first (Lambert, 1975).

Whereas the European Commission promotes language learning and declares multilingualism as an asset, respect for the existing linguistic diversity is expressed by translating all parliamentary documents into all the official languages of the European Union. In the Common European Framework of Reference for Languages (CEFR), where language descriptors and levels for learning, teaching and assessment of languages are described, a distinction is made between multilingualism and plurilingualism. In the CEFR, the latter is advocated: "plurilingualism has grown in importance in the Council of Europe's approach to language learning. Plurilingualism differs from multilingualism, which is the knowledge of a number of languages, or the co-existence of different languages in a given society" (2011, p. 4). The CEFR further explains that multilingualism can be attained by encouraging students to learn more than one foreign language or reducing the dominant position of English. Plurilingualism is defined as building communicative competence that does not keep languages or cultures strictly separated in mental compartments. Instead, the communicative approach means flexible use of all previous knowledge and experience of language, letting languages interrelate and interact—a definition that resembles that of translanguaging, which is discussed below.

Underlying the meaning of language and language policy are ontological, ideological and political conceptions. The recent term translanguaging, which was translated into English from Welsh in 2001, sees language as a dynamic repertoire and not a system with socially and politically defined boundaries

(García & Wei, 2018). Whereas the initial use of the term translanguaging referred to teaching strategies (Williams, 1994), García later transferred the term to the speakers' use of language as a semiotic system or communicative repertoire rather than any established or named languages. Using a translanguaging lens, "languages" are seen as products of nation-state constructs. García and Wei (2018) define a translanguaging perspective as disrupting a view of bi-/multilingualism, noting that "monolingual norms are not natural and they have been socially constructed (Heller, 1999) to exert power as a means of what Foucault [...] calls governmentality" (García & Wei, 2018, p. 2). García and Wei thus promote a dynamic rather than additive bilingualism, referring to humans' agency in choosing how to use language rather than adhering to constructed language norms, thereby putting the student at the center, not the named language. Whereas translanguaging and plurilingualism suggest a disaggregated ontology of languages as meaning making features used by humans, the ontological stance underlying the concepts of mono-, bi- and multilingualism infers a view of language as a fixed system:

Traditional ontologies of language rooted in Saussurean structuralism and Chomskyan generativism are essentialist, that is, they conceive of language as a thing, a bounded object. These essentialist ontologies first divide language (or the linguistic sign) into form and meaning and subsequently accord less or more central weight to form (in Chomskyan perspective) or to meaning (in Saussurian perspective). (Ortega, 2018, p. 69)

Ortega (2018) defines two possible ontological perspectives on language: *essentialist* or *non-essentialist*. The first represents a monolingual worldview with developmental benchmarks to illustrate a native speaker norm and fixed developmental language dimensions or stages. The non-essentialist view is closely related to an empowering ideology with a reluctance to speak of deficit language, but preferably of varieties. In the latter, language is seen as a process rather than a product. Ortega explicitly questions the essentialist perspective, noting that "it is fair to say that the research on world Englishes and the sociolinguistics of globalization has made strict (essentialized) dichotomies along the axis of '(non)nateness' untenable when it comes to research practices" (2018, p. 73). The fluid, dynamic, non-essentialist views on language have become synonymous with the multilingual turn in applied linguistics (Conteh & Meier, 2014; May, 2013). Schissel et al. (2019) refer to a speech given by Li Wei in 2019, where the multilingual turn is turned into the post-multilingual era. Both Li Wei and MacSwan (2017) suggest that the use of the construct of multilingualism equals buying into the ontology of named languages. Within applied linguistics, a marked shift away from the named languages characterizes the discourse surrounding translanguaging practices. Shohamy (2011) even claims that these views follow notions advanced by Makoni and Pennycook (2006) of "disinventing languages."

However, contrasting theories (Busch, 2012) maintain that the traditional language categories cannot be ignored, opposing a dynamic linguistic view of language. By referring to Thorne and Lantolf (2006), Ortega (2018) maintains that SLA (second language acquisition) researchers are unaccustomed to reflecting on the meta-theoretical issue of language ontology and how it affects the disciplinary prioritization of language development. In applied linguistics, the discussion is now concerned with how this shift in language view affects assessment:

[U]nderstanding dynamic models of language as measurable skills poses extreme challenges for the field. It may not be possible to develop language assessments using the same methods, procedures, and concepts that have been used within the field for other forms of language assessments that have been aligned with named varieties. That is to say, constructs based on the idea that languages are bounded and separate entities within an individual speaker would no longer be adequate conceptually and operationally. (Schissel et al., 2019, p. 374)

An understanding of the ontological origin is crucial in discerning how mono-, bi-, multi-, pluri- and translingual stances influence policy, pedagogy, assessment and, in this case, even research. My ontology as a researcher leans on an essentialist stance as outlined above, seeing languages as more or less fixed systems; however, as expressed in systemic functional linguistics (SFL) (Halliday, 2014), different registers may appear in dialogue or be invented and constructed to fit the task or requirements of the situation. A systemic functional grammar connects form and meaning in contexts of use: “every language is a meaning-making system that varies in its expression according to contexts of use” (Schleppegrell, 2020, p. 18). For assessment, this implies that predefined varieties, including subject-specific ways of using language, can be modelled, taught and targeted. However, all varieties cannot be predefined, as language also emerges in use to create meaning. I see all languages as valuable assets that can be used in a translanguaging interim strategy toward the acquisition of the target language, but also together with other translinguals, as a means to an end, not an end in itself.

Language policy

The study of language policy emerged out of a variety of disciplines, including linguistics (applied and sociolinguistics), sociology and political theory. Kaplan and Baldauf (1997, p. xi) posit that language policy is a “body of ideas, laws, regulations, rules and practices intended to achieve the planned language change in society, group or system.” Spolsky and Shohamy (2000) differentiate between explicit language policy, language ideology and language practices to draw attention to the ideological orientation behind conceptualizations of language. Kaplan and Baldauf (1997) further point to differences in how certain policies represent informal statements of good feeling rather than actual policy. They make a distinction between substantive and symbolic, where the first is articulating specific steps to be taken and explores the reality of what specific kinds of changes would entail. The second is concerned with direction and intent, positing a belief in the general goodness of a process. Kaplan and Baldauf (2007, p. 124) bring attention to the importance of knowing who the actors are:

[T]he mid-twentieth-century belief in the adequacy of a broadly based and ideologically neutral approach is no longer enough. We need to know who the planners are and what impact their views might have.

Ahn (2007, p. 4) refers to Kroon’s (2005) “language policy cube,” a tool used for “a theoretical and apolitical” language policy analysis (see Figure 1.4 below):

Because language policies are interventions at some level to change the linguistic behavior of some population, when using the language policy cube, the z-axis needs to be determined first. At what geo-political locale is the language policy being formulated? Then, what language is the policy geared towards (y-axis)? Finally, what language planning domains are involved (x-axis)?

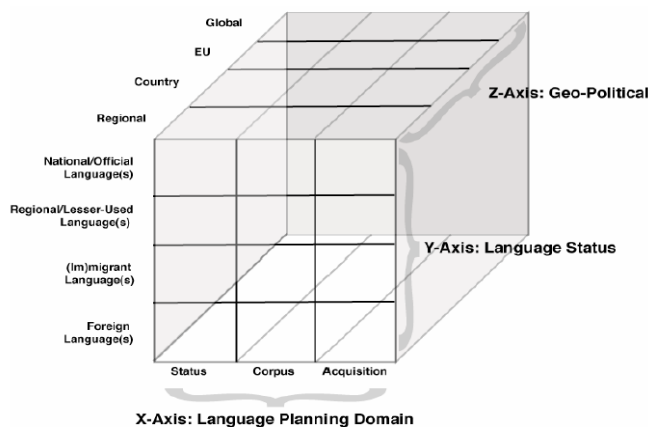


Figure 1.4 *The language policy cube*
Adapted from Kroon (Ahn, 2007).

Without going into detail about how the cube can be used as a tool in policy research, it is worth noting how language status planning involves focusing on the status or position of a language in relation to other languages; this is a relevant issue in relation to the two educational contexts at stake in the present study: CLIL, where English is used as a medium of instruction and migrant education, where Swedish and mother tongue instruction are at the fore. In a similar vein, as stated by Kaplan and Baldauf, Ahn (2007) notes that it is of interest to reflect on who—what agent or stakeholder—is planning for which language in what target domains. Moreover, what is the norm and which register serves as the standard reference?

Shohamy (2003) makes a distinction between *language policy* and *language education policy*, claiming that language policy is concerned with decisions about language use in society, whereas the latter refers to how these decisions are carried out in schools. Menken (2008, p. 402) infers that “language education policy therefore pertains to which language(s) will be the medium of instruction in schools, which languages will be taught, how they will be taught and by whom they will be taught.” Initiatives can be seen in many countries to purposively foster bilingualism by using English (a foreign language) as the medium of instruction; this is often referred to as EMI or CLIL. At universities in Sweden, for instance, English is used more and more frequently, especially in the natural sciences (see the chapter on education in Sweden below). Using English as a medium of instruction in higher education is seen as a measure to welcome international students into the programs (Dalberg, 2013; SOU, 2000:47; SOU 2018:78). International profiles are also offered in certain secondary schools in Sweden, thus creating opportunities for multilingualism while at the same time contributing to segregation between student groups, according to some (Brandén & Bygren, 2018; Saminathen et al., 2018). However, today “perhaps more children and students worldwide are educated in a second language (L2) than in their first language (L1)” (Nikula et al., 2016, p. xiii). Looking at the cube in the figure above, the status of different languages is a factor when determining which language should be used when and where.

Using a foreign or second language as the language of instruction has implications for the students’ ability to access, understand, process and communicate subject content. It also has implications for

academic achievement and assessment, a challenge facing more and more teachers at various educational levels globally. Menken (2008) maintains that high-stakes tests are de facto language education policies that are very powerful in shaping classroom practice, but the policies embedded in the exams are implicit rather than explicit. Referring to the No Child Left Behind (NCLB) legislation that was passed in the States in 2001, she exemplifies how the assessments, which are usually in English only, place English language Learners (ELLs) at a disadvantage, since all exams rely heavily on English language proficiency. Menken argues that high-stakes state assessments offered in English only, together with the repeal of the Bilingual Education Act in the States after NCLB, have wash-back effects on instruction, encouraging English-only approaches. The CEFR (2011, p. 5) acknowledges a paradigm shift from a focus on native-like mastery to plurilingualism, a shift that causes a certain ambiguity since it is claimed to still be important to discern qualifications associated with certain proficiency levels:

The responsibilities of educational authorities, qualifying examining bodies and teachers cannot simply be confined to the attainment of a given level of proficiency in a particular language at a particular moment in time, important though that undoubtedly is. [...] The full implications of such a paradigm shift have yet to be worked out and translated into action.

Students' use of an L2 concerns subject teachers in both settings referred to above, those purposively using English as a medium of instruction, but maybe particularly teachers of immigrant students, since the newcomers have not made an active choice. The challenges need to be addressed, but this also entails a need to align policy, curricula, instruction and assessment with an awareness of how language is viewed in society, how languages are treated and are to be used. Does the norm entail the fostering of mono-, bi-, multi- or translanguals? The answer to that question is ontological and ideological in character and has implications for educational policy and pedagogy in terms of which language needs to be taught and learned, and by whom. Apart from the nation-state categorizations of languages as "legitimated by the state and institutionalized within civil society, usually to the exclusion of other languages" (May, 2001), there are linguistic registers within the languages associated with a target domain and academic discipline, as exemplified below.

The academic language(s)

The importance of language at various segments in education, and in society at large, is incontestable, as already discussed. Language, as we know, serves as a communication tool and is a key to participation and citizenship. Or, put differently, no language means having no way to exchange experiences and knowledge or share life. However, language is not only a means for communicating and mediating between people; it is also necessary for every individual in building knowledge in all subjects (Cummins, 2000; Schleppegrell, 2004, 2005, 2011). The indispensable role of language in the meaning making process is old knowledge, acknowledged both in philosophy—as in the linguistic turn (cf. Wittgenstein)—and in sociocultural theory on learning (cf. Vygotskij), but the role of language in education has been emphasized more and more to date, partly due to more linguistically-diverse classrooms. The recent surge in texts discussing themes such as language across the curriculum, language in all subjects and language development in the classroom bears witness to this (e.g. Bjerregaard & Kindenberg, 2015; Kindenberg & Wiksten, 2017; National Agency for Education, 2012b, 2018e, 2012c). Language is no longer a concern only in language arts (e.g., EFL and SSL) classes, since it is crucial to learning in all subjects.

Language in education can be categorized and described in various ways, the main reason being its many different functions and forms. Academic language represents the language that students need to learn and use to participate and engage in meaningful ways in the content area. Nagy and Townsend (2012) defines it as a specialized language that facilitates communication and thinking about disciplinary content. In curricula standards are expressed using academic language functions (e.g. the ability to critically assess and draw conclusions) to identify verbal behaviors that can serve as indicators that learners have reached a particular goal. Dalton-Puffer (2016) notes that these participant verbalizations make learning intersubjectively accessible.

Cummins coined the concepts *basic interpersonal communicative skills* (BICS) and *cognitive academic language proficiency* (CALP) in relation to the amount of time it takes English language learners to acquire different registers of the language (Cummins, 1984). Whereas a distinction often is made between everyday language and academic language or the language of schooling, the latter can also be broken down into general academic language versus subject-specific language, or content-compatible versus content-obligatory language, which was identified by Snow, Met, and Genesee (1989); see examples in the table below.

Table 1.2 *Content-obligatory versus content-compatible language of schooling*

Content-obligatory (CO) language	Content-compatible (CC) language
<ul style="list-style-type: none"> • Primary, usually learnt first • Necessary to learn the key concepts • Content- or discipline specific • What-oriented, as in the content • Supportive of the “big idea” or “essential understanding” • Required for success with the assessment 	<ul style="list-style-type: none"> • Expands the language beyond academic forms and functions • Provides extra language, or “filler” • Includes more communicative forms and functions • How-oriented, enhancement • Complement and supplement the CO-language

Adapted after Fortune and Tedick. Language objectives for the CBI-classroom (Fortune & Tedick, n.d.)

In contrast to content-obligatory language, which refers to content standards, Fortune and Tedick (n.d.) signal a connection between content-compatible language and language standards in syllabi, thereby implying that there are indeed language standards in the subject matter syllabi. The integrated language goals thus turn all content teachers into language teachers (Gottlieb, 2016). Today, when schools are more linguistically diverse, Met (1997, p. 36) posits that subject content teachers need language learning strategies:

Like all language teachers, they must understand how language develops and be familiar with current pedagogical practices in language education. [...] Over the years we have found that teaching content in a language in which students have limited proficiency differs significantly from teaching that same content in a student’s first language. Teachers need a repertoire of strategies to ensure that students develop both content and language skills.

Although school subjects share common patterns of language use across the curriculum, language in the subject classrooms implies very specific “dialects” of the general academic variety of the dominant language of schooling (Beacco et al., 2016). Theens (2019) expresses that language plays an important role in aiding students in their mathematical thinking and understanding of content, and the use of specific mathematical language and mathematical communication is an important aspect in mastering the subject (e.g., Lithner et al., 2010; Niss & Højgaard, 2011). Geijerstam (2006) notes that students’ writing in the natural sciences differs markedly from that in the social sciences. Students in her study produced fewer and shorter texts in the natural sciences, and these texts contained poorer language. Research also shows that the differences in the disciplinary forms of expression between the natural and social sciences become more apparent higher up in education (Areskoug et al., 2013) and that students find it more difficult to access the disciplinary culture in the natural sciences (Aikenhead, 2005). Schleppegrell (2005) identifies five different features that distinguish formal academic language from everyday language and make it more challenging: density and abstraction; multi-semiotic nature; organization of information in writing; technicality; and the expectancy of a certain stance or authoritative voice. In order to access knowledge, since meaning is constructed in language, Schleppegrell points to the importance of giving the students opportunities to engage with the language, “unpacking” the language to make it visible. Consequently, all students are entitled to instructional practices where the required general language competences are fostered, as well as specific academic subject literacies, whether it be in a first or second language.

The notion of “language knowledge” takes different meanings associated with learning foreign languages, second languages, language by immersion, heritage languages, languages of immigrants and the language of “trans-nationals,” according to Shohamy (2008, p. xiv):

Knowing the English language, the current world’s lingua franca, is different from knowing other languages. Similarly, the language of classrooms and schools may be different from that of the workplaces or communities where bi- or multi-lingual patterns are the norm. Each of these contexts may require different and varied theories of language knowledge and hence different definitions, applications and methods of measuring these proficiencies.

In the quote above, Shohamy identifies an existing language tension, where the need for different named languages, such as English as opposed to other languages, but also specific language proficiency and skills, are dependent on context. The language requirements of different settings do not only imply varying theories of language, but also views on how and when different languages are to be used. This has, or should have, implications for assessment. In the Common European Framework of Reference (CEFR) (2011), international language standards as well as language domains are described. The latter are identified as four separate areas of language use: the personal, the public, the occupational and the educational domains. The CEFR is constructed to accommodate the various needs and forms of language and expresses a desire to be comprehensive and “specify as full a range of language knowledge, skills and use as possible” without pretending to cover all possible uses of language in all situations a priori, which is impossible (Council of Europe, 2011, p. 7). For learning and assessment purposes, levels and descriptors have been identified for six proficiency levels, from basic to advanced language user.

CEF [Common European Framework] should differentiate the various dimensions in which language proficiency is described and provide a series of reference points (levels or steps) by which progress in learning can be calibrated. It should be borne in mind that the development of communicative proficiency involves other dimensions than the strictly linguistic (e.g. sociocultural awareness, imaginative experience, affective relations, learning to learn, etc.). (Council of Europe, 2011, p. 7)

The communicative language competence is the focus of the CEFR, where competences are defined as the sum of knowledge, skills and characteristics that allow a person to perform actions (Council of Europe, 2011, p. 9). While trying to differentiate language in progress, creating scales and descriptors, the CEFR stresses that language learning is not linear or identical for every individual. Learning a language is a process that includes many levels of language proficiency toward a target language norm or functional target language. The basic language levels when acquiring a foreign language can be transferred into denoting the acquisition of a subject language as well. The “language in progress” is sometimes labelled *interlanguage* or *hybrid language*, and involves the mixing of everyday concepts with more specific subject literacies (Gibbons, 2018; Olander & Ingerman, 2011). Gottlieb (2016) labels it *interim language*. In second language acquisition, the term *translanguaging* (García, 2008, 2012) has become more prevalent recently, particularly in guidelines for the teaching of migrant students (see Part 3). If and how translanguaging can be used in assessment with monolingual learning goals and curricula still needs to be settled.

Language and assessment in the subject disciplines

Almost all syllabi express both interdisciplinary language functions and subject-specific language goals, since these are needed in order for students to acquire and use the subject content (National Agency for Education, 2012c). However, assessments in the social and natural sciences build partly on different constructs (i.e., what it means to “know” something or have the required competences or skills). Disciplinary traditions and epistemologies are embedded in the subject-specific literacies. As mentioned previously, each subject has its own “dialect” as well as a shared language of schooling. Referring to Bernstein’s (1999) two conceptualizations of disciplinary knowledge structures, the hierarchical and the horizontal, Airey (2012) describes how the knowledge structure impacts not only the acquisition of knowledge but also the progression in language proficiency. The hierarchical structure seen in the natural sciences implies integrating new knowledge with previous knowledge in a corporate body of expression, whereas in the humanities, new perspectives necessitate new distinct descriptive languages, resulting in a more horizontal progression. Using a disciplinary knowledge structure lens when approaching the integration of language and content affects how language progression is perceived, either as incorporating more and new phenomena into the same explanatory system, as in the natural sciences, or adding isolated new “registers,” as in the humanities.

Besides differing knowledge structures, the disciplines also have their own perceptions of how disciplinary knowledge should be understood and used. The TIMSS science framework, established by the International Association for the Evaluation of Educational Achievement (IEA, 2019) defines the scientific construct that is assessed in the large international assessments using three levels—knowing, applying and reasoning—which represent a hierarchical knowledge structure similar to the one found in Bloom’s taxonomy (Anderson & Krathwohl, 2001). According to the TIMSS framework (IEA, 2019, p. 54) “[s]cientists engage in scientific inquiry by following key science practices that enable them to investigate the natural world and answer questions about it. Students of science must become proficient

at these practices to develop an understanding of how the scientific enterprise is conducted.” Understanding and knowledge acquisition thus require the use of distinct science practices, which are expressed as asking questions based on observations, generating evidence, working with data, answering research questions and making an argument from evidence.

In the social sciences, there are no international assessment frameworks defining the disciplinary construct, but on a Swedish university’s website, it is stated that the historical discipline should be used in different situations. Historical thinking is described as including various historical skills that need to be practiced at school. Stearns et al. (2000) note that assessment of historical knowledge cannot be done using an interchangeable device to convey historical knowledge, since understanding means that students must engage and play a role in “making historical knowledge” (Stearns et al., 2000, p. 3). Comparing the natural and social sciences thus indicates that in the natural sciences it is possible to learn certain standard practices, but the social sciences—at least history—require a more flexible form of expression to communicate knowledge and understanding. Schleppegrell (2005), referring to Martin (1989, 2002), distinguishes that whereas history has been characterized by interpretation and making authoritative evaluations of events, the natural sciences imply technicality in the construction of theories about the world. She concludes that teachers need to understand the linguistic features of their disciplines to help their students succeed.

The language-learning expectations of the subject matter teachers have consequences for assessment of the intended learning outcomes in the subjects. The verbal expressions are communicative actions—Bloom (1956) called them *cognitive domain verbs*—but they are also referred to as *academic language functions* or *cognitive discourse functions* (CDFs). Dalton-Puffer (2016) uses the latter, basing her choice on the speaker’s or user’s communicative intentions and needs rather than taking a curricular perspective, as in previous approaches. Referring to numerous studies and projects with the aim of cataloguing, systematizing and making visible the linguistic requirements in connection with subject content education, Dalton-Puffer expresses a need to synthesize surveys of all verbs extracted from curricular documents to make them more accessible. She extracts seven basic communicative intentions and explains how the CDFs verbalize subject-specific cognitive actions; see Table 1.3 below for examples.

Table 1.3 CDF (*cognitive discourse functions*) types and their members

Function word	Examples
Classify	Compare, contrast, match, structure, categorize
Define	Identify, characterize
Describe	Label, identify, name, specify
Evaluate	Judge, argue, justify, critique, recommend, comment, reflect
Explain	Reason, express cause/effect, draw conclusions
Explore	Hypothesize, speculate, predict, guess, estimate, take other perspectives
Report	Inform, recount, narrate, present, summarize, relate

Adapted and synthesized from Dalton-Puffer (2016).

In the present study, Bloom's revised taxonomy was used for the analysis of the assessment samples in Study I. In Study II, Dalton-Puffer's list of CDFs was used together with Bloom's verbal functions when developing the questionnaire, and Bloom's learning domains in combination with the verbal functions were used in the analyses of the assessment samples (see analytical models below and the method and materials section).

In this study academic language is defined as *subject specific language* (CO language) and *general academic language* (CC language). The first is broken down into subject specific concepts, subject specific skills, as expressed by the use of function words (CDFs), and subject specific genre. The second represents what was described previously as a formal authoritative and cross-disciplinary language (Schleppegrell, 2005). Disciplinary concepts are sometimes seen as an integrated part of content in subject areas (Ball et al., 2016), but here they are regarded as language elements, based on the teachers' own, however sometimes, wavering perceptions (they claim to focus on concepts as vocabulary, but not teaching language).

The focus of the present study is English (Study I) and subject disciplines within the natural and social sciences (Studies I and II):

The natural sciences:

- Biology
- Chemistry
- Physics

The social sciences:

- Geography
- History
- Religion
- Social studies/Civics

The individual subjects represent specific features, but academically they share ontological and epistemological concerns (B. Andersson et al., 2005; Areskoug et al., 2013; Öberg, 2019). Both fields have been characterized as “language dense” and therefore more difficult for language learners than mathematics, for instance (SOU 2017:54). The construct to be assessed and what it means to “know” or be competent in the subjects include content areas as well as language, expressed in cognitive and communicative actions.

Thus far, matters of *what* to assess have been the focus, but in relation to language performance, the *how* of assessment is of crucial importance. Moreover, the *what* and the *how* are not really separable. The *how* can relate both to purpose and format. As regards purpose, assessment is commonly characterized as *formative* and *summative*, or as assessment *for* learning and assessment *of* learning. Although the terms often are dichotomized and the latter—summative assessment or assessment of learning—charged with negative emotions of performance anxiety, the two are not mutually exchangeable, since they serve different purposes as an integral part of the teaching and learning process. Biggs and Tang (1999, p. 195 f) state that both are intended to reflect how well students are doing or have recently done, to “match performance as it is, with performance as it should be.” However, in the assessment of learning, the results are used to grade students at the end of a course or a module. Formative feedback is inseparable from teaching and Biggs and Tang call it a “powerful teaching/learning activity.” Lately, assessment is more often described as assessment *as* learning, seen as an aspect of formative assessment (assessment for learning). The OECD (2013, p. 140) claims that “assessment is a process that helps focus attention toward what matters most in education, beyond just access and participation: the actual learning outcomes of each student”; in so doing, it sets the learner at the center, informing students, teachers and stakeholders about educational progress and student needs.

In the field of language assessment, theories and practices have always been closely related to definitions of language proficiency. Since language is a tool for assessment across disciplines in multilingual schools, experiences from language testing may be of significance for assessment in the disciplines. Shohamy (2008) notes that matching the *how* of testing with the *what* of language reveals several different periods in the development of the field. Referring to different assessment eras, Shohamy describes a development from discrete-point testing, implying objective procedures where language was viewed as consisting of lexical and structural items, to an integrative era assessing discoursal language, on to the communicative era, replicating authentic oral and written interactions, to the performance testing era, where language users were expected to perform real life tasks. In authentic real-world performance assessments, interlanguage variation becomes a concern, as topics and individual performances may differ. Instead of measuring accuracy or grammaticality, “the testing situation shares features in common with particular target language use situations [...] authenticity measures the learner’s ability to function in a specified real-world target situation” (Tarone, 1998, p. 83).

Spolsky (1977) divided language testing into three major periods—the pre-scientific, the psychometric and the socio-linguistic—implying a preference for different methods and formats at different times. No systematized training in assessment or test development was required during the first period. The reliability of assessment was questioned due to the use of open-ended questions graded by different instructors, hence the name “pre-scientific” (Malone, 2008; Spolsky, 1995). “By contrast, the second, psychometric period, emphasized statistics and measurement and moved away from open-ended test questions to test items focusing on discrete aspects of language” (Malone, 2008, p. 227). In the socio-linguistic period, new forms of assessment emerged as a response to the criticism regarding the reliability of open-ended constructed response assessment. In a similar vein, Shohamy (2011, p. 420) posits that the discourse about a so-called multilingual construct (in contrast with one that is monolingual) should be reflected in examples of how language should be addressed in assessment, especially given the context of increasing migration and globalization; this aspect will be discussed below.

Assessment of language learners

Quality in assessment concerns issues of fairness, accuracy, equity and validity—issues which the assessment of language learners brings to the fore. The critical question is therefore if the same standards and methods can be used, since conditions differ. Two other critical questions relate to the purpose and the consequences of assessment. Shohamy (2011, p. 419) questions whether it is legitimate to compare students with different language backgrounds, since they are not being assessed under the same conditions. She further comments that there seems to be a lack of coordination between the two disciplines of teaching and testing, referring to how translingual teaching and learning approaches are currently taking place within the language education field, while the assessment field continues to view language as a monolingual, homogenous, and often still native-like construct.

Standards are used to set up comparable goals; assessment is not only about the individual, but about participants in society, making people functional and possibly employable. Assessment among language learners raises questions as to how to attain predefined standards, intended learning goals or knowledge requirements when the same conditions do not apply. Abedi (2008) infers that there is a substantial performance gap between students for whom the language of assessment is a second language and those who are native speakers, particularly in subjects with high academic language demands. Increasing diversity can be found globally. Abell and Siegel (2011) note that U.S. science teachers are not equipped

to reach all students, especially English language learners (Johnson, 2006; Lee et al., 2007). They maintain that equitable and fair assessment must include assessment tasks that support learning for all types of learners, regardless of their language ability or cultural/ethnic/racial background (Siegel, 2007; Siegel et al., 2008).

In a context with language learners, culturally responsive assessments that take the students' backgrounds into account are especially important. The conscious integration of and attention to content and academic languages, as well as documenting students' progress in relation to both, are advocated, especially since today's assessments must reflect the performance of students, including more targeted use of language and content (Gottlieb, 2016). The use of multiple assessments over time, as in performance assessments that allow flexibility in when and how, while at the same time making progress visible, is advocated, involving group as well as individual performance on tasks (Lachat, 1999), and using multiple methods and a variety of response types (Pitoniak et al., 2009). For accessibility purposes, Educational Testing Service (ETS) recommends using school-based words and not referring to home context, due to diverse life experiences. Activating students' higher-order thinking skills and setting high goals is often mentioned (Bjerregaard & Kindenberg, 2015; Kindenberg & Wiksten, 2017; Lachat, 1999). Gottlieb mentions how the use of culturally sensitive assessment implies screening the students' backgrounds, using the home language initially and helping the students towards academic language development, this being the stepping stone to academic achievement, "the glue that cements content and language learning" (2016, p. 42). Bjerregaard and Kindenberg (2015) and Kindenberg and Wiksten (2017) point to the importance of communicating the intended learning outcomes to make them visible to the students by using recurrent feedback, especially in relation to language goals and teaching the students to give feedback to peers. Defining the construct to be assessed and being attentive to constructive alignment, and planning classroom activities and assessment after defining intended learning outcomes are therefore of utmost importance.

Gottlieb (2016) notes that academic language development involves the use of different genres and covers three dimensions—discourse, sentences, words and expressions—all of which include the language domains of listening, reading, writing and speaking. She mentions how the academic language functions in syllabi can involve "complementary standards" both for language and for content, which can be targeted in the classroom and assessed in performance tasks. In conclusion, Gottlieb argues that setting up language standards in combination with content standards, focusing on academic language in class, differentiation of language according to proficiency level when designing assessment tasks, and teachers' documentation of language progress can serve as a bridge to academic equity. Both Gottlieb and Schleppegrell (2001, 2005) propose that teachers' awareness of their students' levels of language proficiency is essential for successful learning and assessment.

Lately, the increase in older migrant students entering educational settings in Sweden has emphasized the need to address challenges in relation to academic achievement in all subjects. A crucial question is how large groups of older students, with only a few years left to graduate, will attain enough academic language proficiency in the limited time they have to earn their grades in the subject disciplines. Is it possible to ensure fair and valid assessment of language learners' subject content knowledge without risking that insufficient language competence becomes a hindrance? Moreover, do all students get the same opportunities to acquire necessary academic language? And how do teachers perceive their role as language teachers? Shohamy (2011) questions the use of high-stakes national tests with immigrant students, since these tests often require students to be like natives and be measured on monolingual tests in the dominant language, which she concludes results in hierarchies of immigrants and natives being created and perpetuated. Referring to research, Shohamy notes that transfers from the migrants' L1s

continue to be manifest for a long time—maybe a lifetime—something that is considered a liability. Pointing to the construct used, she infers that national test developers, contracted by central governments, help maintain a monolingual policy. Instead, Shohamy insists that the construct needs to reflect the multilingual reality, requiring the adoption of an organic approach where teaching and assessment are inseparable.

Why assessment? What is the purpose? When is it necessary or appropriate to assess, and what are the consequences? These are common questions in relation to assessment, together with what and how to assess. While the *what* refers to the construct to be assessed (the content, the domain or specific skill), the *how* relates to the method or procedure. “The ‘how’ [...] is derived mostly from the field of testing and measurement which has, over the years, developed a broad body of theories, research, techniques and practices about testing and assessment.” (Shohamy, 2008, p. xiii). In the section above, referring to different eras in language testing, the reliability of open-ended language testing formats was questioned, thus leading to psychometric methods. Validity, as will be discussed below, is connected to the consequences of assessment. For the sake of second-language learners, a multiple-choice (MC) format that manages to measure higher-order thinking without having to produce a lot of language could be useful. As Scully (2017) notes, it is a general perception that the MC format is incapable of assessing cognitive processes beyond recall or recognition of knowledge, which is associated with lower-order thinking. According to Scully, MC items should not be considered a homogenous entity, but can be developed to discriminate and measure more advanced cognitive levels. However, constructing good MC items is time consuming and challenging. Moreover, the high-quality distractors required to show higher-order thinking skills involve more of a reading load (Scully, 2017), thus making the questions more language-dense.

As can be seen above, assessment of language learners’ subject content knowledge and language, entails what can be perceived as dichotomous views on assessment, involving either a monolingual stance or a translingual. The latter, if consequently pursued, proposes changing the monolingual construct of assessment as seen in syllabi and policy documents, letting go of standards and a specific target language. The former, on the other hand, implies supporting the performance of language learners by using scaffolding or accommodations, something which will be discussed below.

Accommodations and scaffolding

For language learners, the most significant concern when assessed on subject content is accessibility associated with the nature of the language used in the assessment. Accessible language involves test directions, the vocabulary, syntax and sentence length and structure of the test items, the use of passive voice and cultural references (Young et al., 2014). “Since almost all assessments measure language proficiency to some degree [language learners] may receive lower scores on content area assessments administered in [the target language] than they would if they took the same test in a language in which they were proficient” (Pitoniak et al., 2009, p. 2).

In order to reduce the negative impact of insufficient language, accommodations are sometimes used. Kieffer, Lesaux, Rivera and Francis (2009) list seven types of accommodations:

- the use of simplified language
- L1 dictionaries/glossaries
- bilingual dictionaries or glossaries
- tests in the native language
- dual language tests
- some dual language passages/questions
- extra time

Kieffer et al. (2009, p. 1175) mention that each of these is theoretically justifiable for ELLs as they are designed to address the language needs of the ELLs by minimizing variations in scores because of construct-irrelevant language abilities. Gibbons (2009) speaks of scaffolding rather than accommodations, stressing the need to provide students with support through planning and instruction. She uses a model—often cited among Swedish SSL teachers—called *the genre-based curriculum cycle*, or *the teaching-learning cycle, TLC* (Gibbons, 2002). The model covers four steps: building knowledge of the field; modeling the text type; joint construction; and independent student writing.

Other guidelines (Elmeroth, 2017; Frankfurt International School, n.d.) include provision of an alternative task or modification of the task, as well as offering various types of accommodations. Alternative assessments are exemplified as the use of portfolios to map developing knowledge and skills, but also nonverbal strategies such as role play, when applicable. Further accommodation suggestions include sentence starters or frames (thus reducing the amount of language the language learners must produce themselves) or receiving help in explaining the question or words orally, and teaching and discussing model answers and discussing assessment rubrics prior to the test. Hence, support can be provided in both test construction and in test administration. Sireci and Faulkner-Bond (2015) distinguish between direct linguistic support—e.g., simplifications or translations—and indirect support—e.g., extended time. Referring to research, they note that it is important to not use any accommodations but the ones the students need, making the test more accessible without altering the construct. According to Young (2009), results on tests seem to differ most in questions with high language complexity. Also referring to research, he notes that access to wordlists for instance, has been proven more effective than oral test directions, whereas extended time seems to have modest or no effect. However, there is some evidence that bilingual dictionaries or wordlists in the target language may impact the construct. Abedi (2008) mentions that there have been concerns over validity as a result of certain accommodations. Studies have shown that extended time and read-aloud affect the performance of both second-language learners and native speakers, and therefore this is a threat to the validity of test results. Accommodations provided on tests should match those provided for instruction; otherwise, assessment results may be invalid. Translation of assessment tools into students' native languages should therefore only be used when students have been instructed in that language, so that instruction and assessment are aligned (Abedi, 2008; Sireci & Faulkner-Bond, 2015). Concerns about the validity of an assessment thus relate to whether accommodations can level the field for disadvantaged groups without generating unfair advantages compared with other groups. Stobart (2005) mentions three areas where equity should be considered in relation to fairness in assessment across groups: questions of access (differences in resources); curriculum (what is taught, why, and how); and assessment (appropriateness of form, content and mode for different groups). As mentioned in the introduction, fairness is considered to be closely related to the validity of assessment—a matter that is covered next.

Validity in assessment

Validity is one of the most important attributes of assessment and test quality (Educational Testing Service, 2014; Pitoniak et al., 2009). It describes an assessment's quality in relation to successful function and results. Definitions and conceptualizations of validity have evolved over time. The traditional definition of validity refers to "how well a test measures what it purports to measure," focusing on the test. Validity has further been divided into different types based on various attributes and functions, such as the following (definitions synthesized from the literature):

Construct validity: Does the test accurately measure underlying theoretical non-observable concepts, constructs or traits?

Content validity: Does the test adequately and representatively sample the content area to be measured?

Face validity: Does the test appear valid and relevant to examinees, personnel who administers it and other observers?

Criterion validity: How well do the test scores correspond/show a relationship with a criterion that is the indicator of the ability?

Predictive validity: How well does the test predict later performance on a related criterion?

Content validity has been important in educational assessment, where test specifications often are presented in a table that indicates the content areas and cognitive skills measured (Sireci & Sukin, 2013). Face validity is the weakest form of validity since it is subjective and not a technical term; i.e., a test that appears to be valid is not necessarily so. Predictive validity is often classified as a subcategory of criterion validity, together with concurrent validity. As the interest in consequences associated with test results increased, criterion-related validity was considered insufficient to support claims for validity. More comprehensive views on validity required more evidence than could be generated from the test itself, evidence that was consistent with theories underlying the construct and the intended purposes, while at the same time avoiding unintended negative consequences for individuals (Sireci & Sukin, 2013). By extension, the term "construct validity" now incorporates not only the test, but its proposed interpretations and uses (Kane, 2006) and is thus defined as a unified concept, incorporating other types of validity. Cronbach (1971) defined validity in terms of interpretations and range of potential uses, thereby inferring that validity is not an inherent property of a test, but should be evaluated each time a test is used.

The APA Handbook of Testing and Assessment in Psychology (2013) defines validity as the appropriateness and usefulness of a test for a particular purpose and concludes that the unitary concept of validity incorporates the necessary qualities:

Two concepts are central to understanding contemporary validity theory and test validation. The first is the unitary conceptualization of validity centered on construct validation; the second is the argument based approach to validation. (Sireci & Sukin, 2013, p. 63)

Messick (1989, p. 13) defines validity as "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences

and actions based on the test.” Chalhoub-Deville (2016) notes that Messick’s consequential validity approach has been questioned lately: while consequences are acknowledged as relevant to testing practices, disagreement prevails on whether they should be part of validity. In response to concerns regarding whether consequences should be part of validity, Kane (2006) proposes a framework where justification of test use leans on interpretation/use arguments. In the same vein, Bachman and Palmer (2010) developed an assessment use argument framework for language tests. The model builds on Toulmin’s (2003) structure of practical reasoning and includes data, claims, warrants, backing and rebuttals with the purpose of providing arguments for the quality of test properties and test use. Kane (2013) maintains that consequential research is necessary but not sufficient to support the quality of assessment results. The assessment standards of the American Educational Research Association (AERA) (1999, p. 16) are in line with Messick’s view that evidence about consequences may be directly relevant to validity when traced to sources of invalidity. Evidence that reflects valid differences in performance can be used to inform policy decisions but are not technical aspects of validity. AERA mentions two sources of potential invalidity: two threats to making valid inferences from test results identified by Messick as *construct under representation* and *construct-irrelevant variance*. The latter is particularly relevant in relation to language in a multilingual environment. How can adequate and appropriate inferences be made about a student’s subject matter knowledge and competence if the test contains language elements (i.e., irrelevant features) that the student has not been able to acquire before the assessment? At worst, the test may become a test of reading comprehension or spelling instead of the intended content area. Assessments inevitably contain certain language skills: multiple-choice test items include certain reading skills and constructed response items require writing. No matter which competences are being tested, the students must be given a chance to practice and learn those skills before taking a test. Validity in test results is reduced if a test has sampled a domain in a way that benefits a particular group, since the inferences may be misleading (Gipps & Stobart, 2009). In the same vein, a teacher needs to identify language-related skills and include those in the teaching, since they are part of a content and language integrated intended learning outcome (see p. 22f on assessment of language learners).

An assessment task can consistently measure the wrong thing and thus be reliable but not valid. The validation of an assessment task starts by defining the purpose and ends with evidence bearing on the appropriateness of the interpretations made of the students’ underlying ability based on test scores (Sireci & Sukin, 2013). The validation process takes both evidence and threats to valid use of assessments into account. Crooks, Kane and Cohen (1996) developed the “chain model” for evaluating validity in educational assessment by distinguishing possible threats to eight linked stages in assessment. “Assessment is depicted as divided into eight conceptually distinct stages, with validation then based on careful scrutiny of each of these stages [...] with weakness of any one link weakening the chain as a whole” (1996, p. 266). Crooks et al. posit that the chain model is useful both in the planning of assessments and the development of assessment tasks.

Threats related to the administration of tests include if the students have been given too little time, received inappropriate coaching, had access to resources not normally permitted, or the reverse. “A student may be described as poor at mathematics when the real problem is that the student was unable to read the instructions for the mathematics tasks” (Crooks et al., 1996, p. 271). This can be described as construct-irrelevant variance, which is also a threat to the aggregation of results if inappropriate weight is given to different aspects of performance. If there are substantial differences between the educational goals that are included in the assessed domain and those in the target domain in syllabi, the extrapolation of results outside of the assessment to expected results on the universal target domain is

threatened, and construct underrepresentation may be at stake. Crooks et al. (1996, p. 275) use the example of the too narrow assessment format (the *how*) and too narrow content (the *what*): “if all tasks have been presented as multiple-choice items, the results may not give a sound indication of performance on similar curriculum goals assessed through different task formats.” In relation to scoring, they go on to say:

[S]corers of history tasks who place considerable weight on the correctness of students’ written expression (spelling, grammar, etc.) might be doing an injustice to students whose knowledge of history and skill in historical analysis are strong, but who are poorly equipped to write well in English. (Crooks et al., 1996, p. 272)

In a context with migrant students, it can be challenging for teachers when they almost need to employ editors to improve and standardize the students’ writing (Shohamy, 2011). In order to see through “incorrect or deficit” language, a categorization that is not entirely politically correct to use, both teachers and students need training to decode or develop language skills, respectively: “At first glance constructed responses from ELLs may be confusing to read and may appear to be off topic or unscorable. Many of these responses, however, can be scored [...] if the scorer has been trained to identify and properly evaluate ways an examinee might approach an item” (Pitoniak et al., 2009, p. 19). Another solution is to adopt a multilingual assessment policy, allowing the use of the students’ L1. (The alternatives will be discussed in relation to the findings in the present study, see the synthesis of results, Part 4).

The APA handbook makes reference to what is labelled a relatively new aspect of validity evidence in educational testing: alignment. Alignment is based on how well test content in educational tests aligns with curricula. Equitable assessment requires that teachers assess students in ways consistent with how students are taught. Referring to Webb (1997, p. 4), alignment is described as “the degree to which expectations [i.e., standards] and assessments are in agreement and serve in conjunction with one another to guide the system toward students learning what they are expected to know and do” (Sireci & Sukin, 2013, p. 66). Below, alignment is outlined in relation to teachers’ beliefs and decision making when planning and executing their teaching and assessment practices.

Teacher beliefs and conceptions

Teachers’ beliefs, practices and attitudes are important for understanding and improving educational processes. They are closely linked to teachers’ strategies for coping with challenges in their daily professional life and to their general well-being, and they shape students’ learning environment and influence student motivation and achievement. Furthermore, they can be expected to mediate the effects of job-related policies – such as changes in curricula for teachers’ initial education or professional development – on student learning. (OECD, 2009a, p. 89)

One chapter in the OECD Teaching and Learning International Survey (TALIS) report from 2009 is dedicated to “teaching practices, teachers’ beliefs and attitudes.” The survey investigates teachers’ and school leaders’ self-reported beliefs regarding how teachers apply their knowledge and skills in their teaching practices. Numerous studies note the importance and impact of teacher beliefs on teaching practices (Fives & Gill, 2014). Three sets of beliefs appear essential to teaching practice: beliefs about teaching, knowledge (epistemic beliefs), and beliefs about students’ ability (Fives & Buehl, 2016).

However, there are also studies suggesting that there are inconsistencies between professed beliefs and observed practices (Pajares, 1992) that must be kept in mind. The inadequateness of beliefs as an indicator of behaviors is often associated with the role of contextual factors and the interaction and interconnectedness between an individual and the surrounding world (Li, 2013; Poulson et al., 2001; Tamimy, 2015). The use of data collection methods such as surveys and interviews, as in the present study, involves the participants' subjective statements, where they may say not what they actually do, but what they sense is expected. At the same time, their willingness to share their experiences must be highly valued and appreciated.

Teacher beliefs is but one prevailing term used to define what goes on in teachers' minds; other related terminology includes, for instance, *conceptions*, *perceptions*, *values* and *cognition*. Consequently, it is worth noting the epistemological background of the terms and their associated academic traditions. Entwistle, Skinner, Entwistle and Orr (2000) state that the term *conception* has been used mainly in higher education, whereas the school-based literature in Europe focuses more on teachers' own way of thinking, which is defined as beliefs. In North America, they maintain, the term *conceptions* is used by researchers when describing different aspects of teaching. However, Pajares (1992) notes that in the school-based literature there is little agreement, and therefore the terms beliefs and conceptions are used interchangeably.

In order to attempt to make a distinction between the terms beliefs and conceptions, Entwistle et al. (2000, p. 5) mention beliefs as often reflecting strongly felt ideas and implicit or tacit knowledge that is derived from experience and driven by emotions rather than from any conceptual framework. Beliefs have been argued to reflect personal truths, alternative perspectives to experience reality, affective and evaluative components, guided by feelings rather than logic (Nespor, 1987). Conceptions are claimed to be formed more consciously and to be aimed at distinct concepts with commonly-defined meanings. Leaning on cognitive psychology, the term is defined as extracting the common features of experiences in which the concepts are exemplified. Referring to Sugrue (1997), Entwistle et al. note that teacher perceptions are influenced by teachers' prior beliefs, whereas Barnes et al. (2015, p. 285) note that "the concept of conception subsumes knowledge and belief into a singular construct and provides a framework for describing teachers' overall perception and awareness." In order to make a distinction based on the discussion above, conceptions lean on more commonly agreed upon features, whereas beliefs are imbued with more personal experiences and affective components, and therefore they can be considered complementary.

Years of teaching experience seems to be linked to teaching practices and general conceptions about teaching and learning (J.-L. Berger et al., 2018). However, for teaching practices to be changed, there must also be change in otherwise incompatible teacher knowledge and beliefs. Reconceptualizing assessment practices may give rise to theoretical and practical conflicts in relation to underlying assumptions of assessment and teachers' pre-existing beliefs about teaching and learning (Black & Wiliam, 1998; Matese et al., 2002; Shephard, 2000).

Sugrue (1997) found that student teachers' perceptions of "good teaching," their lay theories of teaching, and their emerging identities as teachers had all been substantially influenced by their prior experiences and beliefs. Using the definition of beliefs above, this may explain findings in research that preservice teachers rely on emotions and experiences from their own schooling (N. Barnes et al., 2015; Graham, 2005) that may contribute to perpetuating traditional teaching and assessment methods. Also, teachers' beliefs and practices in general are expected to be associated with their background characteristics and especially with their professional education (OECD, 2009a, p. 113).

The term “teacher cognition” evolved as a result of the influence of cognitive psychology on educational research during the 1970s, comprising a new focus on teachers’ agency. Teachers were not mechanical robots who simply implemented external prescriptions such as curricula designed by others, but rather exerted agency in the classroom as active thinking decision-makers (Borg, 2003). Borg describes how accompanying terminology such as beliefs and knowledge emerged to support investigations into teacher cognition and subtypes of teacher knowledge appeared, as in Shulman’s (1986) notions of *pedagogical knowledge* versus *pedagogical content knowledge*. Teacher cognition has been particularly influential in the study of foreign and second-language teachers’ practices.

Researchers in the field of educational assessment use varied terminology: hence, teachers’ assessment beliefs, conceptions and perceptions are used interchangeably (Barnes et al., 2015; Brown, 2004, 2006; Harris & Brown, 2009; Remesal, 2007). Teacher cognition, that is often used for teachers’ beliefs in language education, was used in Study I; in Study II, the terms beliefs and conceptions are used, in agreement with research on assessment.

Teacher beliefs and practices are influenced both by personal and external factors that shape the individual practice in a co-constructed manner and setting. Research has been concerned with understanding what goes on in teachers’ minds and what thus affects teachers’ practices. Shulman (1986, p. 8), for instance, formulated questions in relation to teacher learning: “What are the sources of teacher knowledge? What does a teacher know and when did he or she come to know it?” The concept of knowledge can be a subject of discussion, making a distinction between what is scientific and objective versus subjective. For the purpose of this study, knowledge comprises both personal and shared understandings acquired through experience or education, thus leading to beliefs and conceptions. Fives and Buehl (2016), as mentioned above, for example, conceptualize knowledge as epistemic beliefs. The OECD (2009a) uses “professional competence” as an umbrella term for knowledge and beliefs.

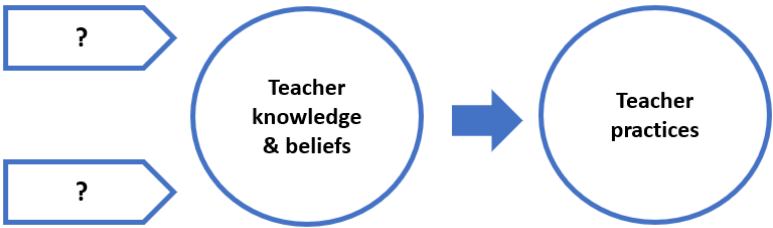


Figure 1.5 What are the factors that influence teachers' knowledge and beliefs, and hence practice?

The question marks at the left in Figure 1.5 above, indicating what feeds into teachers’ beliefs and building teacher knowledge, can derive from external factors sometimes beyond teachers’ control or personal factors. External factors that dictate teachers’ educational context to various degrees are represented by the students, the organization of the school, policy documents and the character of the subject taught, which were mentioned above as epistemic beliefs and thus integrate intersecting external knowledge requirements with teachers’ personal experiences. These factors could be described as frame factors and are not subject to beliefs, but they are also filtered through teachers’ lenses, where previous experiences matter.

The OECD (2009a) points to professional background factors that influence teacher beliefs and practices: *type of training*, *certification* and *professional development*, which would be described as personal factors using this distinction. Above, epistemic knowledge was mentioned, along with teachers’ experiences from their own schooling. In research, personal socioeconomic aspects are included as possible features for comparison. In order to synthesize the factors mentioned above, Figure 1.6 below gives an overview of the key factors influencing teachers’ practice. The figure is not exhaustive but sets out to accommodate the most important aspects. Other than type of training, which includes teacher education and professional development, there are three more categories. The second is named “subject identity,” as disciplinary tradition influences epistemological beliefs and therefore pedagogy. This also covers disciplinary knowledge requirements as seen in policy documents. The third, “professional experience,” includes years of teaching and other work-related practice, but also education level (elementary, secondary) and experiences of different student groups—in this case, language learners and migrant students in particular. Shulman (1986) believed in merging the two fields of teachers’ subject content knowledge and pedagogical method into pedagogical content knowledge (PCK).

Pedagogical content knowledge [...] includes an understanding of what makes the learning of specific topics easy or difficult: the conceptions and preconceptions that students of different ages and backgrounds bring with them to the learning of those most frequently taught topics and lessons. (Shulman, 1986, p. 9)

In the figure below, the third factor represents Shulman’s PCK the best; however, boundaries are not clear cut and the category subject identity incorporates certain content knowledge. The fourth factor, “personal background,” covers other personal factors that may contribute to shaping teachers’ beliefs (e.g., experiences from own schooling, age, gender, cultural differences, and experiences from being a language learner, to mention some that are of interest in the current study). Figure 1.6 below tries to capture some of these sources.

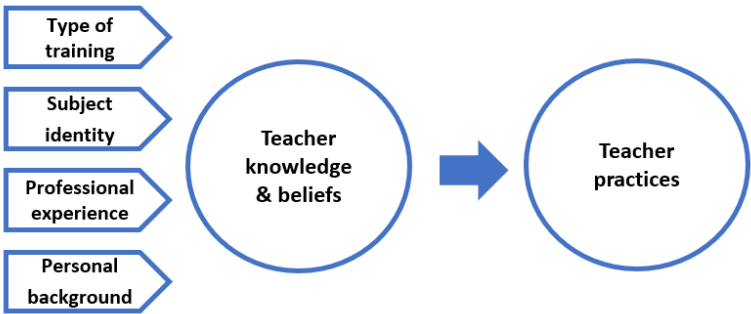


Figure 1.6 Background factors influencing teacher beliefs

As mentioned above, boundaries are “leaky” and aspects that are integrated into the categories in the figure above have been separated as individual independent variables in the questionnaire and the analyses in the present study (see Study II in particular). What is important is how teachers’ knowledge and beliefs affect the students. Each teacher holds a set of beliefs that determines his/her views as to how students acquire knowledge and thus determines their priorities in practice. Teachers’ experiences

and perceptions of the students’ abilities and capacity to develop language skills to show their understanding of subject matters will influence how teachers plan their instruction, as mentioned above. Biggs and Tang (1999, p. 27) note that “desirable student learning depends both on student-based factors—ability, appropriate prior knowledge, clearly accessible new knowledge—and on the teaching context, which includes teacher responsibility, informed decision making and good management.” Young (2013) states that teaching depends on the knowledge teachers have of their subject, the individual students and how they learn, and the knowledge that informs what teachers require their students to do. Young (2013) further highlights the difference between school knowledge and everyday knowledge, since they are constituted by concepts that are different in both structure and purpose. Everyday concepts, in a similar vein as experiences, are context-specific:

[W]ithout the opportunity to engage with the concepts of a subject-based curriculum, children’s understandings are inevitably limited to those contexts and those experiences. In contrast, the concepts associated with a subject-based curriculum are not tied to specific contexts; they are linked to each other and the underlying theories associated with the subject in question and underpinned by the community of subject specialists. (M. Young, 2013, p. 110)

Young thus points to the importance of building bridges between students’ everyday experiences and curriculum content to accommodate students who lack the everyday experiences that would allow them to generalize when faced with subject-specific concepts (something that can be considered an important role of education) in order to compensate for students’ diverse backgrounds. Students’ experiences vary substantially, but they are crucial learning resources and therefore teachers need to take student background into consideration when planning instruction. Figure 1.7 below is an adaptation of Bigg’s constructive alignment model, aligning instruction with the intended learning outcomes (ILOs). The figure also incorporates the “student factor” (the blue box in the middle).

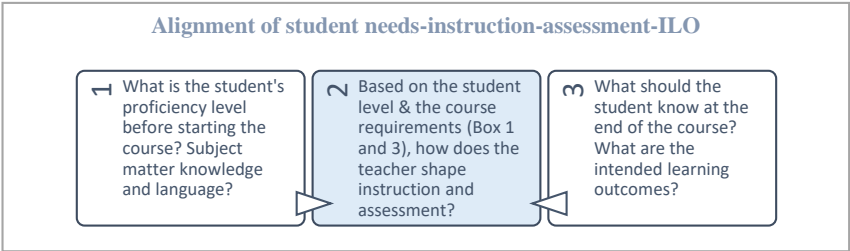


Figure 1.7 Adaptation of the model of constructive alignment

Biggs (2003) suggests that the intended learning outcomes should be formulated first to achieve well-planned instruction and teaching, with learning activities designed second and assessment third. The research questions in this study focus on what happens in Box 2 in the figure above. However, in order to understand what shapes teachers’ assessment practices, it is important to distinguish teachers’ perceptions of the intended learning outcomes and goals (Box 3). Of equal importance is to understand how teachers perceive the students (Box 1). Even though alignment is directed toward the intended

learning outcomes at the end of a course, as in Box 3 in the figure, the arrows in Figure 1.7 point toward the middle (as this is the focus of the study), hopefully leading to student attainment, in Box 3.

In summary, teacher beliefs and conceptions, building on individual as well as external factors, have an impact on practice, including assessment practices:

Teacher beliefs are of paramount importance in determining how the ideas they wrestle with at the knowledge base level are interpreted and put into practice. The TALiP model encourages the inclusion of teacher beliefs on cognitive and emotional dimensions in the development of assessment literacy building. (White, 2018, p. 7)

The TALiP model referred to above stands for Teacher Assessment Literacy in Practice, and was developed by Xu and Brown (2016) by combining two fields of research: educational assessment and teacher education. Xu and Brown argue that student teachers' responses to assessment theories is affected by whether the principles are consistent or dissonant with their current conceptions. Below, teachers' professional knowledge and skills in relation to assessment in particular are discussed.

Assessment literacy

The term assessment literacy has been conceptualized as a set of beliefs, knowledge and practices about assessment that guide teachers to use assessment in order to improve student learning and achievement (MAC, 2017). All teachers need professional knowledge and competence concerning the purpose and consequences of assessment. Popham (2009b) notes that many teachers and other test users have a limited understanding of assessment fundamentals such as reliability and validity, and thus limited assessment literacy.

Abell and Siegel (2011) developed a model to describe secondary teachers' assessment literacy in diverse science classrooms among language learners. The model includes four knowledge elements: *knowledge of assessment purposes*, *knowledge of what to assess*, *knowledge of assessment strategies* and *knowledge of assessment interpretation and action-taking*. All four aspects are based on teachers' views of learning in combination with assessment values and principles. The model has been adapted for this study and is described below in the section on analytical models (cf. p.40f). The last knowledge element concerning interpretation can be directly translated into concerns for consequential validity (Messick, 1989). Popham equates assessment literacy with an understanding of the consequential basis of assessment:

[I]f a teacher mistakenly believes that validity resides in the test itself, the teacher will be inclined to defer to whatever results the "valid test" produces. Assessment-literate educators, however, understand that education tests merely provide evidence that enables people to make judgmentally based inferences about students. (2006, p. 84)

Referring to Shulman's (1986) concept, pedagogical content knowledge, and Magnusson et al. (1999), Abel and Sigel (2011) conclude that knowledge of assessment in science include two dimensions: knowledge of what to assess and how to assess (cf. p.21 and 24). These two dimensions provide a foundation for assessment literacy. The first dimension is closely tied to instructional goals and the second to teachers' knowledge of different methods of assessment. Drawing on the *assessment triangle* proposed by the U.S. National Research Council (Pellegrino et al., 2001), Abell and Siegel (2011, p. 207) add three complementary dimensions for assessment literacy: a model of students' cognition and

learning in the domain; a set of beliefs about the kinds of observations needed to provide evidence of the students' competencies; and an interpretation process for making sense of the evidence. These three dimensions cover the following three components: the student, the assessment format and the interpretation of outcome.

All teachers need to develop assessments appropriate to the contexts where they teach, and balance is needed between different measures, including formative assessment strategies. Popham (2009a, p. 7) maintains that teachers who are genuinely assessment literate both know how to create more suitable assessments and are familiar with "a wide array of potential assessment options." Popham (2004) goes so far as to describe assessment *illiteracy* as professional suicide.

Literature on teacher assessment literacy in science is limited, as are studies examining teacher knowledge of assessment directly, although there is rich literature on classroom assessment (Corrigan et al., 2011). In a review of the research on science teacher knowledge, Abell (2007) found only a few studies in which researchers attempted to examine science teacher knowledge of assessment theory and skills directly. Malone (2013) mentions that one issue when trying to define good assessment practices is contingent upon testing and teaching experts' differing ideas regarding aspects of assessment literacy. For the purposes of the present study, a model is used that operationalizes assessment literacy as being able to respond to questions in relation to why, what, how, and the consequences of assessment based on informed assessment principles and values; this is illustrated below in Figure 1.12 and serves as one of the analytical tools presented in the following.

Analytical models

Several tools have been used for the analyses of the qualitative data in the two studies, as outlined in the methodology chapters. This section provides a description of the models in order to explain the background of the concepts applied in the study.

The common denominator in all models is their relationship to assessment and/or language in policy and in pedagogy. The objective was to find tools to render teacher assessment beliefs and practices in the context of language learners visible. Table 1.4 below provides an overview, naming the analytical tools and showing when and where they have been used before going into more detail about their individual features.

Table 1.4 Analytical models used in relation to different parts of the study

	Study I	Study II	Synthesis
Cummins’ matrix and academic function words/CDFs*	X	X	X
The content-language integration continuum	X		X
The assessment literacy model		X	X
The assessment continuum		X	X
The language integration matrix and profiles		X	X
Curriculum alignment	X	X	X
Assessment validation	X	X	X

*Academic function words in the knowledge requirements and test items were used in combination with Cummins’ matrix.

Some models were more instrumental in one of the two studies and others in both, but all contributed to the results in the synthesis of the two studies. Cummins’ matrix is used particularly in Study I together with the academic function words, alternatively labelled cognitive discourse functions (CDFs) (Dalton-Puffer, 2016). The matrix was particularly useful when comparing levels of cognitive load, linguistic load and context embeddedness of the test items in the assessment samples. The function words were used in analyses of alignment between syllabi and assessment across the different data collection methods.

The content and language integration continuum (Lyster & Ballinger, 2011; Met, 1999) is often seen as an attempt to define the content versus language dominance of the CLIL approach; therefore, this model was used in the initial study. The language integration matrix (Leung & Morton, 2016) and its related profiles were used in Study II and the synthesis of results. The model that served as the source of inspiration had not yet been invented when the first study was conducted.

The analyses of the assessment samples and interview data in Study I prompted the use of validation models inspired by Bachman and Palmer (2010), Crooks et al. (1996) and Kane (2010, 2013). The broader data collection in Study II involved more focus on teacher beliefs, implying a need for tools integrating aspects of teacher assessment literacy; here, I drew on Barnes et al. (2015) and Abell and Siegel (2011). However, the validation perspective has been key throughout the whole process, since validity is essential for equitable assessment. In the present study, the validity evidence was gathered and explored using test content and evaluations of possible testing consequences (cf. Sireci & Faulkner-Bond, 2015); here, these were based on teachers' self-reported assessment procedures and self-developed tests.

Appropriate and aligned instruction is mentioned as key to access and fairness in assessment practices among language minority groups (Gipps, 1994; Stobart, 2005). In the same vein, the literature mentions the importance of backward design in CLIL (Llinares et al., 2012), starting with the intended learning outcomes and the characteristics of the students in order to be able to provide relevant instruction and feedback. Hence, constructive alignment (Biggs, 2003; Porter, 2004) was considered throughout the process and became especially important for the synthesis of results in Part 4.

Not all renditions of results explicitly reflect the use of the analytical tools, as they sometimes were instrumental during the process but not as visible in relation to the results. Below, the analytical and conceptual models are briefly outlined.

Cummins' matrix

Cummins (1981) developed a model to help illustrate the interplay between the degree of cognitive complexity in activities and assignments in school and the amount of context provided as support. Cummins' matrix included two scales (see Figure 1.8 below): one vertical, going from *cognitively demanding* to *cognitively simple*, and one horizontal, going from *context-embedded* to *context-reduced*. The model has been elaborated and presented in several versions, all with the purpose of conceptualizing how assignments in school need to be framed to motivate and help language learners.

Context-embedded activities (to the left) are represented by oral discussion in class, for instance, where the interlocutors can negotiate for meaning and ask for clarifications, whereas context-reduced activities (to the right), such as assignments in a textbook, require knowledge in the language to be able to lean on textual cues for support. Subsequently, Hall (1995) added the features *linguistically less demanding* (to the left) and *linguistically more demanding* (to the right) as used by Coyle (1999), arguing that the challenge for teachers in the CLIL context is to create cognitively demanding tasks while at the same time using less demanding language; in other words, ensuring that language does not get in the way of understanding (see Quadrant A in Figure 1.8).

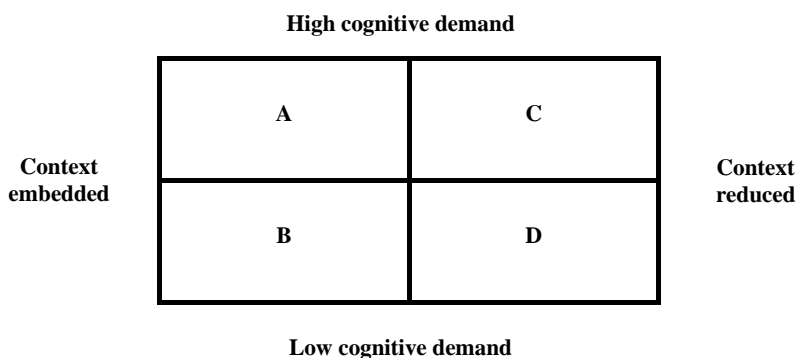


Figure 1.8 Cummins' matrix from 1981

The original version of the matrix has been further developed, creating a more teacher-friendly version according to its originator, Cummins (2017), with the additions of the dimension of *big* versus *small challenge* on the vertical axis and *more* versus *less support* on the horizontal axis (Gibbons, 2009; Mariani, 1997; Wilson & Devereux, 2014). Gibbons (2009) notes that the most favorable setting for multilingual students involves cognitively demanding assignments with a lot of scaffolding (Quadrant A in the matrix, Figure 1.8). Challenging tasks with a lot of scaffolding should be used in all subjects, according to Gibbons, in order for students to learn the language of schooling they need to succeed. Below are some examples of function words (cf. Coyle, 1999) in relation to each quadrant that can be found in assignments:

A: High cognitive demand/challenge, context embedded/concrete, more support, linguistically less demanding

Generalize, summarize, plan, classify, compare, contrast

B: Low cognitive demand/challenge, context embedded/concrete, more support, linguistically less demanding

Identify, name, match, retell, describe observations, apply known procedures

C: High cognitive demand/challenge, context reduced/abstract, less support, linguistically more demanding

Analyze, argue a case, develop ideas, justify opinions, evaluate, form a hypothesis, apply principles

D: Low cognitive demand/challenge, context reduced/abstract, less support, linguistically more demanding

Copy, repeat, reproduce information

The Coyle (1999) matrix, used in CLIL context and slightly differently labelled (1–4 instead of A–D), suggests an order where the low cognitive and linguistic demand in Quadrant D should not even be considered. Instead, the desired progression goes from lower-order thinking skills with a lot of interaction and context (Quadrant B in Figure 1.8) toward higher-order thinking skills, providing contextual support along the way. The goal in CLIL is therefore to move from A to C over a period of time (see Figure 1.9 below) by granting accessibility of language in order to learn both language and content (Coyle, 2005). As regards the design of test items and assessments, it could imply going from multiple-choice items in B toward more open “why-questions” in A and C, where A is providing more contextual scaffolding and C represents the most advanced and elaborated assignments—an argumentative essay, for instance (Cummins, 2017; Hood & Tobutt, 2009).

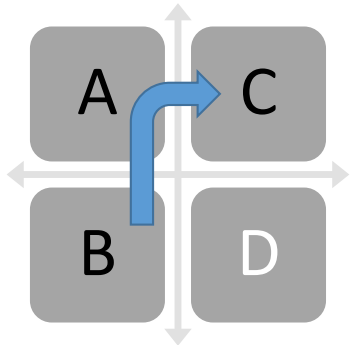


Figure 1.9 *Desired progression in assignments and student proficiency*
Adapted from Cummins (2017), Coyle (1999) and Hood & Tobutt (2009)

Cummins matrix is used for the analysis of the assessment samples in Study I and Study II.

The content-language integration continuum

The content-language integration continuum was developed by Met (1997) and Lyster and Ballinger (2011) to illustrate the variety of integrated approaches, ranging from a primary orientation to language versus a primary focus on content in CLIL-related teaching contexts. A need was discerned to differentiate between approaches where language was added as a second learning goal or as a “bonus” in subject content courses (to the left of the spectrum), versus approaches where subject content was added and used to render language courses more authentic (to the right); variants of intermediary stages between the two were also included. The content-driven approaches are sometimes called strong or hard CLIL, whereas the language-driven are labelled weak or soft CLIL (Ball et al., 2016).

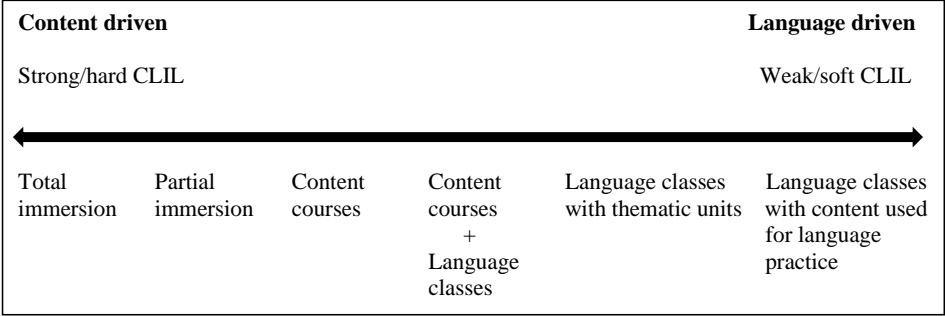


Figure 1.10 The content-language integration continuum
 Model adapted from Lyster & Ballinger, (2011, p. 280) and Met (1997).

The continuum model is used to define the CLIL practices in Study I and for comparison of the Swedish CLIL and NAS teaching contexts in Studies I and II in the synthesis.

The assessment continuum

According to research on teachers’ beliefs and conceptions about assessment, teachers can be categorized along a continuum with more or less fixed positions; teachers driven by pedagogical motives can be found at one end, and teachers driven by pressure for accountability and goal attainment at the other (N. Barnes et al., 2015). A continuum means that teachers and their assessment practices can be found anywhere along the spectrum, but also change positions (see Figure 1.11). Schools are driven by a motivation to reflect how well students are doing in relation to intended learning outcomes at the end of a course and during a course (referred to above as summative and formative assessment purposes), implying there are dual purposes for assessment without one excluding the other.



Figure 1.11 Continuum of teachers' conceptions about assessment

Table 1.5 below gives an account of basic aspects in pedagogy- versus accountability-driven assessment, as found in research and identified by Barnes, Fives and Dacey (2015).

Table 1.5 Features relating to the assessment profiles

Pedagogy-driven	Mix	Accountability-driven
Learning-oriented practices Advances teaching Guides student learning Helps to make instructional decisions	→	← Performance-oriented practices Supports external reporting Shows student performance Helps teacher effectiveness
Assessment for improvement Assessment for learning Promotes learning autonomy	→	← Assessment for accountability Assessment of learning Student accountability Teacher accountability

The continuum and the features in Table 1.5 are used together with the language integration matrix (see below) for the initial analyses of the teacher interviews in Study II. The model helped identify teachers’ conceptions, beliefs and assessment strategies. The features identified by Barnes et al. that are outlined in Table 1.5 also rendered the tension visible in teachers’ assessment mission between making judgments about students’ performance and promoting student learning. Since the teachers work at school levels where they are required to grade the students, they are all accountability-driven by default, but to a greater and lesser degree. However, especially in the context of NAS, the diverse proficiency levels make assessment more or less pedagogy-driven depending on school level, as can be seen in the results.

The assessment continuum is used to conceptualize teachers’ assessment practices and is applied to illustrate the teachers’ primary assessment motives, especially in the synthesis of results.

The assessment literacy model

The assessment literacy model was suggested by Abell and Siegel (2011) and further developed in this study. Abell and Siegel explain that the model attempts to capture the various types of assessment knowledge and skills teachers need. At the center is the teacher’s view of learning, which relates to certain values and principles about learning and assessment in the subject, which in turn guides assessment decision making. The model has been adapted to the needs of the present study, and therefore views of the learner have been added to the model.

Abell and Sigel (2011, p. 211 f) explain that teachers’ values and principles interact with four categories of assessment knowledge, which were briefly described previously in the section on assessment literacy: assessment purposes (why); the construct (what); assessment strategies (how); and assessment interpretation and resulting actions (and). The latter, the “and” was suggested by Erickson and Gustafsson (2017) as one of six “basic questions” to be used when planning assessment (i.e., Why? What? How? When? Who? And?). Erickson and Gustafsson maintain that the challenging question “And?” is of central importance when considering the effect an assessment may have on students and teachers’ assessment practices. Therefore, the question is closely related to consequential validity in assessment.

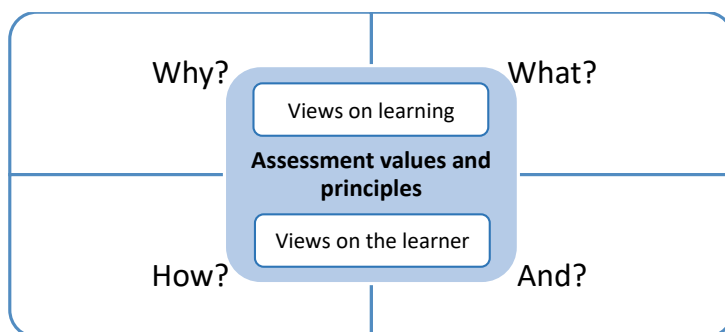


Figure 1.12 *The assessment literacy model*
Adapted from Abell and Siegel (2011).




The assessment literacy model is used in the triangulation of the teacher interviews, survey data and analyses of assessment samples in Study II and for the synthesis with Study I.

The language integration matrix

One model that served to cover and synthesize the teaching approaches in Study II and the synthesis of both studies leans on a matrix developed by Leung and Morton (2016, p. 235 ff). They developed the model while attempting to synthesize the approaches to integration in a number of studies on CLIL teaching and pedagogy (see Table 1.6 below). The matrix combines language visibility in the teaching with language explicitness in policy and curricula. Drawing on Bernstein (2000), the conceptual matrix distinguishes between a *competence* versus a *performance approach*, and *visible* versus *invisible pedagogy*. Competence implies signs of the natural predisposition of the learner, whereas performance refers to requirements to present programmed knowledge and skills. In Bernstein's terms, visible pedagogy is connected to prespecified skills, whereas invisible pedagogy refers to natural competence. Bakhtin's (1981) ideas concerning centripetal and centrifugal tensions in language and culture serve as another conceptual framework for the model. As the words suggest, centripetal implies centralization and unification, whereas centrifugal signals decentralization and diversification.

According to Leung and Morton (2016), Quadrant 1 contains approaches where the development of language competence and meaning combines specific disciplinary content with skills and thinking processes, as in subject literacies. Quadrant 2 represents approaches targeting acquisition of discipline-specific concepts and competences, but not disciplinary-appropriate ways of expression and content-compatible academic language; therefore, this represents less visible language pedagogy. Quadrant 3 implies a cross-disciplinary and functional approach to language competence, combining content and language teaching domains in a general sense, including cognitive discourse functions (CDFs), as seen in many competence-oriented school curricula. Quadrant 4 represents a view on language competence as multilingual, multifaceted and multimodal. Here, the discourse is situated and dialogic, as in translanguaging practices. Structural linguistic features as well as content-language expressions are not the most fundamental consideration for communication; instead, the focus is on pragmatic competence, according to Leung and Morton. Table 1.6 summaries the key features in relation to each of the four quadrants.

Table 1.6 Four different approaches to integration of language
Adapted from Leung & Morton's (2016) analytic framework.

Four different approaches to integration of language and content			Visible language pedagogy	
			More  Less	
			Explicit language-oriented instruction	Incidental language learning
Disciplinary orientation to language	Higher 	Centripetal Centralization and unification	<ul style="list-style-type: none"> • Clear descriptions of linguistic goals. • Specific stages of acquisition. • Language before content. • Focus on explicit subject literacies and skills • Turn function verbs into learning objects. Can-do statements as in the CEFR • Scaffolding to remedy the use of an L2. • Performance oriented. 	<ul style="list-style-type: none"> • Natural acquisition view of L2 (immersion) • Already existing abilities can flourish and grow. • Content before language, language an additional benefit • Focus on discipline-specific concepts and competences. • Language a tool for participation • Competence oriented
	Lower 	Centrifugal Decentralization and diversification	<ul style="list-style-type: none"> • Focus on general literacy norms, less attached to specific disciplines. • More or less functional. • No explicit stages of acquisition. • Lessons seen as speech events using function verbs, sociocultural perspective • Individual choice and agency • Teachers mediating and making accessible 	<ul style="list-style-type: none"> • No strong separation or classification of linguistic varieties, L1 or L2, or subject-specific registers • Language competence is multifaceted, multilingual and multimodal. • Expand communicative repertoires to be able to participate. • Translanguaging the norm. • L1 use not a problem. • Content meaning = created in use, dialogue, not in words.

Leung and Morton (2016) note that the boundaries are “leaky” (illustrated here by the arrows), since issues are influenced by contextual and disciplinary variabilities. The quadrants are “conceptual constructs” and practices are not always consistent. As already noted, the original model was invented to gain an understanding of various approaches to language pedagogy in studies in a CLIL context. As Leung and Morton (2016, p. 237) note, “CLIL issues are at the intersection between principle and practice where conceptual adherence is necessarily influenced by contextual and disciplinary variabilities.” The influence of contextual and disciplinary variabilities is just as much at stake in the teaching of subject content to NAS, as will be seen in the following.

The integration matrix above is used as point of departure to help identify teachers’ approaches to language-in-teaching based on the interview data and the assessment samples. Four profiles emerge, which are positioned in the four squares of the matrix as follows:

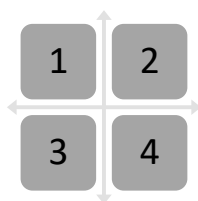


Figure 1.13 Positions of the language integration profiles

In order to illustrate what the four possible combinations of features in the matrix look like when applied to teacher characteristics and practices in the present study, four profiles have been defined, as seen below. Table 1.7 shows which intersecting features are combined in each of the four profiles described below.

Table 1.7 Four possible language integration profiles

	Higher disciplinary orientation to language standard forms of language	Lower disciplinary orientation to language non-standard forms of language	More visible language pedagogy explicit instruction	Less visible language pedagogy incidental learning
Profile 1	X		X	
Profile 2	X			X
Profile 3		X	X	
Profile 4		X		X

Profiles 1 and 2 share *a unitary language ideology*, expressing a goal for students to become proficient in a specific target language that is considered the norm, represented by a national language or other predefined varieties. The target language is expressed in policy documents and can be defined as the disciplinary language and the language of schooling. Language competence is to make meaning in disciplinary-appropriate ways and language processing must precede the learning of content. Bakhtin (1981) argues that a unitary ideology represents a view of language as coherent and implies a pull toward homogeneity, standardization and correctness. Barwell (2014) notes that the standardizing ideology is characterized by ideas of correct spelling and pronunciation. This can be compared with what was defined as a monolingual language policy previously (p. 12f).

Profiles 3 and 4 represent *a diverse language ideology*, accepting more non-standard forms of expression, seeing language as social, dialogic and instantiated in use (Barwell, 2014). The focus is on the development of pragmatic competence, not pre-specified elements. Language is a construct and learning can be seen as a socialization into language, according to a Bakhtinian perspective (Barwell, 2014). Learning is social, not just cognitive. Language is concerned with individual rather than institutional choice. Bakhtin uses the term heteroglossia to describe the tension between different languages and the diversity of voices. Compared with language ideology in policy, these profiles can be identified as disaggregated and translingual.

Profiles 1 and 3 share a view that language should be taught explicitly; they represent a visible language pedagogy, an expression borrowed from Bernstein that means that the teaching is strongly classified and framed. Instruction is form-focused, but without being linked to specific content. Leung and Morton (2016) express that instruction implies a strong focus on form and literacy norms in the general language of instruction.

Profiles 2 and 4 represent a view of language as being learned incidentally, implying an invisible language pedagogy (again in Bernstein’s terms) where instruction is weakly classified and framed. General linguistic and pragmatic competences can develop in this context, according to Leung and Morton (2016). Content teaching is the basic concern, whereas language is an additional benefit.

The interplay between language view and language pedagogy will be discussed more in depth in relation to the results in the synthesis of Studies I and II in Part 4.

The curriculum alignment model

A final model appeared useful for the synthesis of results from Studies I and II. For the purpose of curriculum assessment as well as investigation of curriculum alignment, different “curriculum indicators” are conceptualized. The first, the *intended curriculum*, represents what society envisages as important to learn. This can also be labelled the *planned curriculum*. What is actually delivered in the classroom is described as the *implemented* or *enacted curriculum*, and the third level, what the students learn, as the *attained curriculum* (Porter, 2002). Thijs and van den Akker (2009) and van den Akker (1997) use the terms intended, implemented and attained, which are used for this study. Table 1.8 below summarizes the levels with descriptors adapted from UNESCO (International Bureau of Education, 2016).

Table 1.8 Curriculum indicators

Intended (Planned)	Implemented (Enacted)	Attained (Learned)
A set of formal documents (i.e., national curriculum and syllabi) that specify what the relevant national education authorities and society expect students to learn and develop at school in terms of knowledge, skills, understanding, values and attitudes.	The actual teaching, learning and assessment activities taking place in schools through interaction between learners and teachers, as well as among learners (e.g., how the intended curriculum is translated into practice and actually delivered).	The knowledge, skills, understanding, values and attitudes that learners acquire as a result of teaching and learning, assessed through different means and/or demonstrated in practice. N.B. This may differ from the intended and the implemented curriculum.

The intended curriculum is sometimes defined as macro-level (government), whereas the implemented is characterized as meso-level (school/classroom) and the attained as micro-level (learner). In the model used for this study, the assessed curriculum is placed together with the enacted or implemented, since assessment is considered an integrated part of the activities in class. In some other versions, the assessed curriculum is placed with the intended or the learned.

The alignment model is used in the discussion of results in the synthesis of results in Part 4.

Assessment validation

Validity refers to the use of tests and assessment tasks for particular purposes (cf. p. 26f). Validation is a process to evaluate the degree to which the intended uses can be met. In evaluating the validity of a test for a subgroup such as language learners, the primary concern is test fairness and the measure of fairness is comparability. As Young (2013) notes, academic content tests generally assume close to grade-level language proficiency, which cannot be guaranteed in a CLIL context, and even less so for NAS and their academic language proficiency in Swedish. Validation of assessments includes questions in relation to test development that are based on construct and content validity and the type of linguistic skills and knowledge needed to succeed when taking a test. Validation also includes the administration of assessments; e.g., making them more accessible to subgroups by using various forms of accommodations, without altering the construct (Sireci & Faulkner-Bond, 2015). Questions in relation to construct validity, content validity and test administration were considered in Studies I and II; this was particularly the case in relation to the potential threat of *construct-irrelevant variance* (Messick, 1989), which is based on the need for language skills that may interfere with the test-taker's ability to show knowledge and understanding of subject content. The questions that were instrumental in looking at the assessment samples and teacher reports were connected to the validation chain model developed by Crooks et al. (1996). A sample of questions and features that were explored are presented below; for more information, see Reierstam (2015, p. 43ff):

- *Administration*: Which item format is used? Test layout and instructions? Type of accommodations?
- *Scoring*: What is assessed? Language/content? How is the test graded/assessed? Holistic, analytical assessment? Collegial collaboration?
- *Aggregation*: Balance between item and assignment types for the total score (i.e., selected response vs. constructed response)?
- *Extrapolation*: Relevance between assessment formats in relation to the targeted construct (i.e., the knowledge requirements)?

The validation questions were used in Studies I and II in the analyses of the assessment samples and for the triangulation of results. Note that the questions were used as an analytical tool; for the rendition of results, however, the focus is on the research questions and the broader concept of validity and fairness in access opportunity to show subject content knowledge.

In the following, the contextual background in the Swedish setting is outlined.

Contextual background – The Swedish setting

This chapter provides a background to the Swedish setting to facilitate the understanding of teachers' practices and beliefs in the two teaching contexts in Studies I and II. First, a description of the multilingual landscape in Sweden, the status of different languages, and English in particular is presented, followed by a brief description of the different teaching contexts offered for NAS versus CLIL and an overview of the Swedish education system. Thereafter, a short description of teachers' professional development options is provided, followed by the relationship between language and knowledge requirements in national curricula and policy documents, before a presentation of assessment and grading policy. A comparison of key features and Swedish research in relation to the two contexts—CLIL and NAS—closes the chapter. Together, the aspects provide a backdrop for examining conditions, perspectives and assessment practices in Swedish bi- and multilingual educational settings in CLIL or among newly-arrived migrants.

Multilingualism

Multilingualism in Sweden, as in many other parts of the world, is a result of globalization, migration and international mobility, either by elective choice or by forced circumstances. Linguistic diversity has increased rapidly due to a recent rise in immigrants and asylum seekers, which peaked in 2016. The OECD (2014) noted that Sweden is among the OECD countries with the largest foreign-born population. In a survey from 2016 (Eurostat, 2018), Sweden is ranked the most bilingual country in Europe; 97% of Swedes aged 25 to 64 self-report being able to speak at least two languages. Sweden has a population of slightly over 10 million (Statistics Sweden, 2019) and Swedish is the official language; in 2009, it was reported to be the mother tongue of 8 million (Parkvall, 2016). There are also six recognized minority languages, including sign language, spoken by approximately 2% of the population, according to Parkvall's report.

English is very present in the daily lives of the youth especially through games, music, film and other so-called “extramural” activities (Sundqvist & Sylvén, 2016; Sylvén, 2005). In 2018, Sweden ranked first out of 88 countries and regions in the world's largest ranking of countries by English skills (Economic Policy Institute, 2018). The 2018 EPI report comments that Sweden has returned to first place after a couple of years' absence, even though Sweden has welcomed more refugees per capita over the past five years than any other European country; the inference is that the immigrants' presumably lower English proficiency has had no negative effect. The EPI report (2018, p. 24) concludes that “Scandinavians on the whole have remarkably high levels of English, thanks to strong education systems, daily exposure to English in the media, and an entrenched culture of internationalism.” The status of English goes back in history, as it was already a required and mandatory subject in schools in 1849 at the upper secondary level (Svartvik, 1999). English was considered a vital tool for gaining socioeconomic opportunities (Cabau-Lampa, 1999; Teleman, 2003). Hult (2012, p. 251) notes that “although English has had a presence in Swedish education for at least 160 years, the nature of English in Sweden and how ELT [English language teaching] should be implemented continues to be negotiated”; here, he is referring to the status and role of English in language policy and curricula, where Hult speaks of “tensions.” One such tension appears in compulsory schools using English as a medium of instruction, where it is stipulated by law that 50% of the instruction has to be in Swedish (SFS 2011:185, Ch.9, §18f). This goes back to the Language Act of 2009 (SOU 2008:26), which was passed to protect the Swedish language from domain loss. The role and status of English is thus not undisputed.

In elementary and secondary schools, EMI or CLIL programs have increased, especially in Grades 4 to 9, where independent schools are attracting more students thanks to a free choice system (Hennerdal et al., 2018); the International English School (IES) is one of these. English has become more of a second language than a foreign language, even though it is not, at least for many young people. English is the first foreign language to be introduced in education, sometime between the first and third grades. A second foreign language is taught from the age of 12 or 13 (Grade 6 or 7); historically, this was German and French but in the early 1990s, Spanish was introduced. More languages are eligible in Grade 10, either as an additional language or to substitute for another language. However, the status of English is considered to undermine the motivation to learn other languages (Runblom, 2018); this is a current trend in spite of the European Union's multilingualism policy goals for every young European citizen to know at least two foreign languages (L1 + 2) (European Council, 2002). This goal was formulated and agreed upon in 2002 and in 2014, CLIL was recommended as a method to speed up language learning. The policy goal was re-emphasized in December 2017 (Eurydice, 2017).

English has become more and more common as a medium of instruction in tertiary education as well. Dalberg (2013) notes that there is a substantial increase in English-medium programs, but it is asymmetric and is more prevalent in specific subject domains and universities. The internationalization of tertiary education is part of a conscious and politically-driven incentive for internationalization in order to attract more students from other parts of the world (SOU 2000:92) and for harmonization of higher education within the European Union through the Bologna process after 1999. According to Dalberg's report, the importance of English has grown the most in the social sciences, law, trade and administration, with an increase in courses offered in English, but also in technology and manufacturing. Surprisingly, the increase of EMI in the natural sciences and mathematics is not as prominent, according to Dalberg. Salö's (2010) report may provide some explanatory data, noting that in the natural sciences English is the "high status language" and the number of publications and doctoral theses remains on a consistently high level, with frequencies between 90% and 99% since 1978. In the social sciences and technology, frequencies have gone up significantly. Salö mentions that 65% of advanced courses are given in English, and 9 out of 10 doctoral dissertations are written in English. On the university admissions website (Swedish Council for Higher Education, 2019), it is noted that some universities have formulated their own English language requirements for the EMI courses at the master's level.

In a survey from Stockholm University (Bolton & Kuteeva, 2009), students and lecturers in science, social science, law and the humanities were asked about their attitude to EMI, their exposure to English and the frequency of English usage at the university. A majority of staff and students favored the use of English as a medium of instruction at the postgraduate level, and teachers thought it was important to have some undergraduate courses entirely in English as well. Between 30% to 50% of the students thought it was important to have some courses in English. Strömberg Jämsvi (2019) notes in her doctoral dissertation that multilingualism—beyond Swedish-English bilingualism—is not valued in higher education, nor are minority languages or immigrant languages acknowledged in relation to higher education. She notes that the dominant discourses in relation to language in a Swedish university setting are based on inclusion, success, and marketization. Inclusion in higher education requires sufficient language proficiency in Swedish at the undergraduate level and English at the postgraduate. She further notes that university teachers in all subject courses are expected to be teachers of academic language, but only in English and Swedish.

Looking at languages present in society as a whole, there are several minority languages, especially in the northern parts among the indigenous Sami population, for instance, and the languages of many migrant groups. In a survey from 2012, Parkvall (2016) provides a list of languages in Sweden with the

largest numbers of speakers. He notes that Finnish still is the second most spoken language in Sweden after Swedish, explained in part by minority native speaker groups living in the north of Sweden and by many years of migration and moving back and forth from the neighboring Nordic country. While Finnish still is the second largest language in Sweden with 200,000 speakers, Arabic is closing in with only approximately 50,000 fewer speakers; this figure has possibly changed since 2012 due to migration. The largest languages in 2012 after Finnish and Arabic in due order were “BKS” (Bosnian, Croatian, Serbian, Montenegrin), Kurdish, Polish, Spanish, Persian, Danish, Norwegian, English, Somali and Aramaic. All of these languages had more than 50,000 speakers each.

By 2011, every fifth student in Swedish compulsory school or upper secondary school (see the section below for a description of the education system) had a foreign background, meaning the student was either born in a different country or born in Sweden to at least one foreign parent, according to the NAE (2011). Six years later, in 2017, one-fourth of all students in Swedish schools had a foreign background (National Agency for Education, 2017a). Language and cultural diversity are not evenly distributed, but instead contribute to segregated islands of societal or individual multilingualism. Segregation, possible domain loss, tensions in language policy, and confusion in relation to the status and use of different mediums of instruction are current issues that need to be addressed in relation to equitable education and future development in society.

Education

School is mandatory in Sweden from the age of six, when children start the compulsory preschool class. Compulsory education comprises Grades 0–9 and is organized by municipalities or independent providers (“friskolor” in Swedish), the latter a result of the “free school reform” in 1992. Seventeen percent of the compulsory schools had an independent provider in the 2018/19 school year (the same number as the year before), counting 15% of the students (National Agency for Education, 2018d, 2019a). Almost all students continue to upper secondary education at the age of 16, choosing from 18 national programs, 12 vocational programs and 6 with a focus on university studies (National Agency for Education, 2020a). For students who are not fulfilling the requirements for the national programs there are introductory programs with individually adapted education. They follow syllabi for compulsory school, aiming at reaching the goals for Grade 9 for students to become eligible for a national program or an employment. In 2019, 16% of all students finishing Grade 9 were not qualified for national programs, but for the group of NAS who come when they are older, only 29% qualified (National Agency for Education, 2019e). Entrance into the programs is based on final grade point totals, and students must have a passing grade in Swedish/Swedish as a second language, English and mathematics, in addition to five other subjects for entrance into the vocational programs. For theoretical college preparatory programs, a passing grade in four additional subjects is required.

For NAS there is a language introduction program. The aim of the program is to provide immigrant youth with an education with an emphasis on the Swedish language to prepare them for a national program in upper secondary, other forms of education or work. The home municipality is responsible for providing preparatory education such as in language introduction and other preparatory introductory programs. In 2017, there were reports that language introduction had become the largest program of all in terms of the number of students (Gunnarsson & Helmersson, 2017). The Education Act (SFS 2010:800, Ch.3, §12) stipulates that study counselors must be provided in the students’ strongest language to give the NAS the best conditions for academic attainment and meeting the course requirements. However, the NAS also have the right to an adapted schedule and extended time frame.

Sweden has one of the most decentralized education systems in the world (Nusche et al., 2011), dating back to an administrative reform in 1991. An important consequence of the reform is that municipalities or independent private providers have wide-reaching autonomy, as do the teachers. After the reform, municipalities (instead of the state) became responsible for the allocation of resources and received employer responsibility for the teachers. Education is funded by taxes and as a result of another reform, the school-choice reform in 1992, all children in Sweden, regardless of background, have the financial means via vouchers to attend the school of their choice. Municipalities are required to provide schooling for their residents and municipal schools generally have a local uptake of students, whereas independent schools admit students according to order of enrollment.

Equity in education has recently become an urgent matter of debate in Sweden, as inequalities across students and schools have been widening and children are increasingly segregated into schools with others from similar backgrounds (National Agency for Education, 2018a; OECD, 2019a). At one end of the spectrum, families choose to go to another municipal school rather than the one that is closest to their homes. In suburban areas and major cities, more and more Swedish families choose to enroll in independent schools such as the International English School, an option available almost solely to those born in Sweden due to restricted entry numbers and long waiting lists. In order to be admitted to the International School, families often need to enroll right after a child is born. At the other end of the spectrum, certain areas and towns have welcomed large numbers of newly-arrived migrant students after the big migration wave in 2013–2015, thereby adding to segregation. (See figure below).

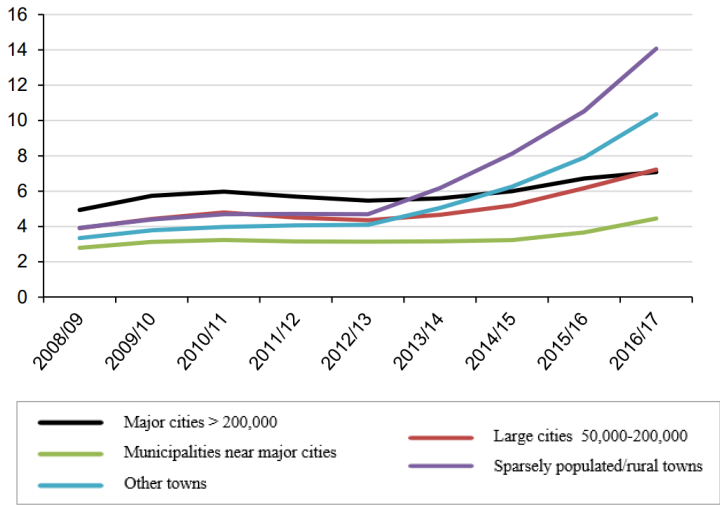


Figure 1.14 Proportion of newly-arrived students, including background unknown, in compulsory school for different municipal units during the schoolyears 2008/09 to 2016/17

Source: (National Agency for Education, 2017b, p. 18)¹

¹ The classification is based on the Swedish Association of Local Authorities and Regions’s (SALAR) categorization of Swedish municipalities from 2011. The revised classification from 2017 is used for this study, see p.109.

In some schools, up to 90% of the students have a foreign background, whereas in others, sometimes in the same town, this applies to only 5% (Dagens samhälle, 2017). In the fall of 2016, 89% of the municipal schools had welcomed NAS, compared to 69% of the independent schools (Hellstén et al., 2020; National Agency for Education, 2017b). The NAE notes that a change has occurred recently: whereas the proportion of NAS in the compulsory schools was larger in the major cities between 2008 and 2012, there was a shift in 2013 with an increase in the proportion of NAS in smaller towns. NAS represent 14% of the student population in smaller towns and the smallest numbers of NAS are found in suburban areas of both major cities and medium-sized towns

The NAE is the central administrative authority for the public school system, publicly organized preschools, school-age childcare and adult education, according to their own website (National Agency for Education, 2020d). Schools are guided by national policy documents; the Education Act is the most powerful steering document for compulsory school. Compulsory school and upper secondary school are also controlled by mandatory ordinances and regulations, such as the Ordinance for Compulsory School (SFS 2011:185), the curriculum, and syllabi with knowledge requirements for the individual courses.

The Swedish national curriculum presents an educational ideology with an appreciation for diversity, while at the same time establishing common goals. See below for some examples from compulsory school (English version); note that the same is expressed in the curriculum for upper secondary school:

The internationalization of Swedish society and increasing cross-border mobility place high demands on the ability of people to live with and appreciate the values inherent in cultural diversity. (National Agency for Education, 2018c, p. 5)

Under the heading *An equivalent education*, it says:

Teaching should be adapted to each pupil's circumstances and needs. It should promote the pupils' further learning and acquisition of knowledge based on pupils' backgrounds, earlier experience, language and knowledge.

The Education Act stipulates that the education provided in each school form and in school-age educare should be equivalent, regardless of where in the country it is provided. National goals specify the norms for equivalence. However, equivalent education does not mean that the education should be the same everywhere or that the resources of the school are to be allocated equally. Account should be taken of the varying circumstances and needs of pupils. There are also different ways of attaining these goals. (National Agency for Education, 2018c, p. 6)

The curriculum thus posits that "equivalent education" is not the same as "equal education" and that there are different ways of reaching the goals. Further down, the curriculum mentions the students' and their guardians' rights to exercise influence and choice, thereby prompting schools to specify their goals, content and working forms for transparency.

Since the enactment of the Home Language Reform in 1977 (Hyltenstam & Milani, 2012), students have had the legal right to mother tongue instruction; by 1968, however, migrant students already had access to "home language" instruction (National Encyclopedia (NE), n.d.). Students with a mother tongue other than Swedish must be offered mother tongue instruction in school, but participation is voluntary. According to the Compulsory School Ordinance, the municipality is obligated to offer mother tongue

instruction, but only if there are at least five students who are entitled to teaching in the language in question and desire instruction, and if an appropriate teacher can be found (National Agency for Education, 2020c). The parents apply for mother tongue instruction or, in upper secondary, the student. As part of mother tongue instruction, NAS may receive study guidance in their mother tongue, which is regulated in the School Ordinance (SFS 2011:185, Ch.5, §4). Mother tongue instruction is considered beneficial for both knowledge and identity development. However, there is considerable variation and lack of equivalence in how the teaching and guidance are conducted and organized (National Agency for Education, 2009).

Teachers and their professional development

In order to qualify as a teacher in Sweden, a minimum of three years and up to five and one- half years of study at university are required, depending on orientation or school level. In 2018, approximately 70% of the teachers in compulsory school were certified and 80% of those in upper secondary school. The NAE notes that the number of qualified teachers has increased slightly in upper secondary, but the reverse is true for compulsory school. It is not illegal to hire a teacher without a teaching diploma, but employment is restricted to a single year at a time. In 2015, the Swedish School Commission was asked to submit a proposal to improve learning outcomes, the quality of teaching and equity in Swedish schools. In the final report, suggestions are made concerning teachers' professional development and a national professional program is proposed. The OECD (2015b) notes that it is challenging to create agreement and national consensus in such a decentralized system. In 2015, the Swedish government gave the NAE an assignment to create so-called "national school development programs" (Department of Education, 2015) to provide professional development for teachers and schools with the intention of providing all students with equal opportunities (Government Offices, 2015). Information and communications technology (ICT) was a special focus area. In 2016, the "appropriation directions" to the NAE added more assignments, one of which was to strengthen the quality of education for NAS.

Looking at the various subject disciplines, the lowest numbers of qualified teachers are found in the subject of Swedish for immigrants (National Agency for Education, 2018d). According to a report from the Institute for Evaluation of Labour market and Education Policy (IFAU) (2014), the aforementioned reform from 1991 has had an effect on relative teacher competence, which has deteriorated compared to other employees, who have in general become more educated since then. Recent reports mention the current challenges of getting teachers to stay in the profession. The OECD (2019a) notes that teaching needs to become more attractive to address teacher shortages, which are most acute in remote areas. In her dissertation, Samuelsson (2019) points to an "era of educational restructuring" with contradictory professional expectations for teachers which she explains by marketization, privatization, governing by results and new public management. Samuelsson calls for increased consciousness of how different contexts for formulation and realization, embedded in different institutional logics, affect teachers' work. She mentions how the preconditions can differ considerably between schools, even within the same municipality, and how teachers respond differently to organizational structures and expectations depending on the local context. In Sweden, professional development for teachers varies depending on local school policy as to whether courses and conferences can be held during working hours or during assigned in-service training days. It may be a matter of negotiating with colleagues to cover for an absence that prevents teachers from attending courses. According to the TALIS survey (National Agency for Education, 2019f), every fourth teacher in Sweden did not believe last year's professional development had any positive effect. The areas in most need of professional development include ICT

as a teaching tool and special needs students. The teachers' self-reported needs for professional development in assessment have gone down considerably since 2013. Teachers in Grades 7–9 express more needs in general than teachers in upper secondary, especially as regards teaching in a multilingual setting.

Language in curricula and policy documents

In Swedish syllabi, language is described as the primary tool for learning. An appended report from the NAE, with commentaries about language and knowledge development in all subjects and at all school levels, establishes that national curricula for the various school levels are based on a conception of the central role of language, to grant all students the opportunities to advance as far as possible (National Agency for Education, 2012c, p. 4). The role of language is mentioned with reference to identity formation, as a tool for learning and as an intended learning goal. In relation to the last goal, Swedish is mentioned explicitly in the curriculum for compulsory school:

Language, learning, and the development of a personal identity are all closely related. By providing a wealth of opportunities for discussion, reading and writing, all pupils should be able to develop their ability to communicate and thus enhance confidence in their own language abilities.

The school is responsible for ensuring that each pupil on completing compulsory school can use the Swedish language, both in speech and writing, in a rich and varied way. (National Agency for Education, 2018c, p. 11)

Among the overall goals in the curriculum for the compulsory school, one of the first goals states that every student should be able to “use the Swedish language in a rich and varied way.” Looking at different ontological perspectives on language, this represents an essentialist view with a Swedish monolingual norm (cf. previous discussion on language ideology, p. 12ff). At the same time, recent policy documents for NAS recommend using the students' first or strongest language to facilitate access to subject content (National Agency for Education, 2018g, 2012c). The same orientation can be seen in the report mentioned above, where an implicit plurilingual approach can be distinguished, making an explicit connection between language and the students' identity formation. A vacillating stance toward language can be distinguished, as both Swedish and students' global linguistic resources are mentioned (translated by the author):

Since Swedish is the majority language in Sweden, the knowledge requirements and intended learning outcomes are first and foremost aimed at the students' development of the Swedish language. However, current research shows the importance of letting children and students use all their available languages as a resource for their learning and development. (National Agency for Education, 2012c, p. 10)

Below are a few samples of instruction strategies mentioned in the guidelines from the NAE (2012c, p. 5) to promote students' language development in all subjects (translated by the author):

- Encourage the students to use all their languages in order to develop these and to advance their subject matter knowledge.
- Apply a language perspective in the teaching and work deliberately with language and content at the same time.

- Analyze the selected content and the knowledge requirements [...] to identify language aspects linked to planning, engagement, evaluation and assessment that the students need to be taught and learn.

The list covers additional aspects of using a dialogic communicative approach for students to learn language through participation. Other examples include providing sample texts and examples of subject-specific conversations to model language. It is further noted that language competence is promoted in authentic communicative situations in all subjects and that school is responsible for supporting students' language development and preparation for higher education and citizenship.

As already noted (cf p.46), it is stated by law that a minimum of one-half of the instruction in compulsory education must be in Swedish (SFS 2011:185, Ch.9, §18f). However, for upper secondary schools in Sweden, no such regulations or even recommendations exist. For upper secondary schools using a CLIL approach, this means that it is up to them to decide how much of the target language they can and should use.

Swedish curricular goals are claimed to be very language-dense. The academic language functions or verbal behaviors found in the learning goals are communicative in nature and usually require language usage. Note the example below from the knowledge requirements for the lowest grade ("E") in the syllabus for Biology 1, the first course in Grade 10 in upper secondary school:

Students use with some certainty the language of science and adapt to some extent their communication to purpose and context.

The need for a disciplinary language is explicit, 'the language of science'. A possible question to ponder is if "use with some certainty" ties in with what was described as an interim hybrid language previously (cf. p.19)? In the aims stated in the syllabi, one of the first is to have knowledge of the disciplinary concepts and use those, as it says in the syllabi for history for instance:

The ability to use different historical theories and concepts to formulate, investigate, explain and draw conclusions about historical issues from different perspectives.

Table 1.9 below provides an overview of some of the most commonly used CDFs in the knowledge requirements for the lowest passing grade (E) in syllabi for the first course in upper secondary, Grade 10. The subject disciplines are the same as those represented in the present study, except for geography, for which there is no official English version of the subject syllabus. The purpose of the table is not to give an exhaustive presentation, but to provide an illustration. Some function words that appear randomly in some of the syllabi have been excluded in the table (e.g., *discuss, express, formulate hypotheses, interpret, illustrate, make simple reflections, use concepts/terms*), since the purpose is to compare disciplines and the focus is on general academic language. Moreover, since only one school level (upper secondary) and the descriptions for the grade E have been used, the sample is small and is only meant to be illustrative. As noted by Dalton-Puffer (2016, p. 33), the structure of the function types are complex, the categories are fluid and their "borders are fuzzy"; translating these from Swedish makes borders and synonyms even more fluid at times. Therefore, the English version was used for this table, which is provided by the NAE only for the upper secondary school level, and not for all subjects, as already noted.

Looking at Bloom’s cognitive descriptors and the levels in Cummins’ matrix (cf. p. 36f), a progression between *lower-order* (this includes Bloom’s categories *knowledge* and *comprehension*) and *higher-order thinking skills* (Bloom’s categories *application*, *analysis* and *evaluation*) can be connected to the function words (Anderson & Krathwohl, 2001; Verma, 2014). However, Bloom et al. (1956) made no specific references to where the lower-order skills end and the higher-order skills start. Scully (2017) suggests that the taxonomy represents a continuum, as seen in Table 1.9 below, where lower-order is found to the left and higher-order to the right. Higher-order thinking skills require more critical thinking and cognitive processing. The main subjects included in the present study are listed on the left in Table 1.9.

Table 1.9 Examples of CDFs and number of appearances in the knowledge requirements for E, the lowest passing grade in Swedish subject syllabi for upper secondary school, Grade 10

	<i>Lower order</i>						<i>Higher order</i>	
	Give an account (report)	Give examples/exemplify	Argue/give arguments/reason	Describe	Explain	Draw conclusions	Analyze	Evaluate/make assessments
Biology	3	-	1	1	-	-	1	3
Chemistry	3	1	2	1	-	-	1	3
Physics	2	1	2	1	-	-	1	3
History	3	3	1	-	2	1	-	1
Religion	5	1	2	2	-	2	4	-
Social Studies	3	-	1	-	1	2	2	1

Based on this small sample, and the objections above about the fluid categories, it can be noted that *give an account* is the most frequently used. It is compared with Dalton-Puffer’s “report” (which she compares with narrating), where real-life events become like storytelling. “Report” can also be substituted with “present,” “recount” and “summarize.” Dalton-Puffer (2016, p. 49) notes that this CDF concerns “what happened, when, who did it, to whom and under what circumstances.” Looking at differences instead, to *evaluate* seems more common in the natural sciences, whereas *analyzing* is more common in two of three social science disciplines. Analyze means “to compare and contrast” according to Bloom, whereas Dalton-Puffer uses “classify” and explains that this is another key CDF, moving away from the observable that can be described toward the more abstract. As regards the word evaluate, Dalton-Puffer (2016, p. 41) notes that it has not been treated extensively in the literature on language for specific purposes. *Describe*, which is mentioned as a key element in academic thinking skills and academic language is surprisingly infrequent in these syllabi: in history and social studies it is not mentioned once; instead, *explain* is used in these subjects, but not in the other disciplines. This may possibly say something about the character and construct of the subjects. However, it should be noted that the categorization of verbs in taxonomies should be done with care, as verbs can be placed in multiple categories and depend on the context (Scully, 2017). Other subject-specific as well as general

academic features need to be considered as well (cf. Table 1.2). As was noted above, only the knowledge requirements for the lowest grade, E, were classified in the table; however, the Swedish syllabi do not use a hierarchical structure for the knowledge requirements. All cognitive process levels are included at each level and progression in the subject disciplines and attainment of more advanced grades is instead identified with the help of descriptors such as “with some certainty” versus “with certainty,” as exemplified on p. 53.

Assessment and grading

Looking at the subjects’ constructs above brings attention to the intended learning outcomes and what should be assessed, but also to the *how* and *why*. The present grading scale in Sweden was introduced in 2011. Grades are assigned on a scale from A to E, A being the highest passing grade and E the lowest; F means fail. “For each subject and course there are knowledge requirements (standards) for grades E, C and A that state what the pupil needs to achieve in order to be awarded a particular grade” (National Agency for Education, 2017c). Whereas the knowledge requirements before 2011 were influenced by the hierarchical view of knowledge as found in Bloom’s taxonomy, for instance, a more non-hierarchical view was expressed in 2011. Hence, progression in the acquisition of knowledge is not solely attached to the cognitive level of the function verb (cf. Table 1.9). The verb *analyze* is therefore not only associated with a higher level of knowing, but seen as a dimension that can be used at all levels together with “modifiers” that distinguish the qualitative differences (e.g., *analyze simple problems* = E, *analyze complex problems* = C, A) (Carlgren, 2013). Grades are assigned starting from sixth grade in compulsory school and teachers are more or less autonomous in making decisions when grading students (National Agency for Education, 2018b). For the sake of assessment, the Education Act (SFS 2010:800, Ch.3, §2) states that all students, regardless of school form, must be given the same opportunities for academic success based on individual competence and educational standards.

Sweden has a fairly extensive national assessment system. The purpose of the national tests is to give all students equitable opportunities to show their skills and support teachers in making fair and valid assessments (National Agency for Education, n.d.-b). National tests are developed by universities by order of the government via the NAE and given annually in certain school years and subjects. The most test-dense school year is Grade 9, when national tests are mandatory and students are tested in Swedish, mathematics, English and one of the subjects in the natural (biology, physics or chemistry) and social sciences (geography, history, religion or social studies), depending on random choice, with schools taking different subjects each year. At upper secondary, the tests in the natural and social sciences are optional, but tests in Swedish, mathematics and English are mandatory. After the education reform in 1992, when private independent schools were allowed public funding, national high-stakes assessment became more important to ensure equity in assessment and admittance into higher education. Yet, equitable grading has proven problematic. According to reports from the School Inspectorate (2017, 2018), as was mentioned in the introduction, schools show discrepancies between students’ results on national tests and the assigned grades. As a result, the government decided to make an addition to the Education Act in 2018, stipulating that results on national tests shall “be especially considered” in grading (SFS 2010:800, Ch.10,§20a).

NAS are to be graded using the same knowledge requirements, but whether the NAS should participate in the national tests or not is up to each school to decide. The NAS can either be exempt or do only certain parts of the tests if schools decide that the prerequisites are not there for the test to provide useful information. Accommodations are possible—e.g., using a mother tongue instructor as translator for

certain words or concepts—but entire tests cannot be translated. The teacher may help explain difficult expressions when the test starts or provide a dictionary for the test in Swedish. The English version of the subject content tests is available only for students in English-speaking schools and is not intended to be used by NAS (National Agency for Education, n.d.-a)

As regards NAS and assessment, the NAE (2018g) expressed in an online professional development module that the format of assessment needs to be adapted to the students' Swedish proficiency, and that initially the students' strongest language should be used. Peer assessment is suggested as a useful learning tool. The teacher is encouraged to give feedback in relation to the language the student needs for the specific assignment, thus focusing on specific language skills. For grading reasons, the NAE posits that an absence of subject concepts does not always equal a lack of knowledge, in the same way that mentions of concepts do not necessarily equal content knowledge.

In 2018, the NAE was issued a government assignment to develop assessment material for initial diagnostic testing of newly-arrived upper secondary students' previous subject knowledge in order to support faster access to subject courses. Since the spring of 2016 it has been mandatory to conduct mapping of NAS's literacy, numeracy and previous educational experiences in compulsory school. The material is available in 14 languages and covers various generic skills in subject syllabi, such as reading and argumentative skills, for instance. In numeracy, mathematical skills are covered. Reports signal that the implementation and use of mapping varies a lot (Nowak Ottosson & Öhlund, 2016).

Research on assessment in the subject disciplines in a Swedish context is limited. In the field of the social sciences, Odenstad (2011) posits that there is a big difference between individual teachers and academic programs in upper secondary school in terms of how assessment is done and what is being assessed in social studies. She further notes that the curriculum does not seem to have a major influence on the content of the tests, but exams rather leans on a kind of disciplinary tradition. Jansson (2011) notes that teachers' assessment practices in social studies are varied, and not focused on the reproduction of facts, as the literature says. The syllabus plays an important role in planning course content, but at the same time the disciplinary tradition and the textbooks drive what is taught in history, according to Casservik (2005). Lindmark (2013) posits that teachers' perceptions of the subject depend on teacher background to a large extent. In a study on students' writing practices in social studies, Lindh (2019) finds that students' writing is strategically and ritually motivated and not communicative in nature, which she attributes to the strong focus on tests, assessment and grading. Lindh maintains that this is explained by the strong focus on assessment and performance culture in the Swedish national curriculum.

In the natural sciences, there is even less research. Anker-Hansen (2015) investigates how to assess scientific participation in civic practices at lower secondary school. He concludes that assessments designed for assessing scientific literacy as situated processes is problematic in combination with the view of scientific knowledge as possessions. For authentic classroom assessment to become useful, he notes that teachers and students need to focus on the quality of participation instead of preconceived ideas or specific scientific theories. In a study on schools' preparations for national tests, Löfgren et al. (2018) distinguish varying enactments of policy—ranging from organized, to systematic to ad hoc—which they conclude have implications for equivalence between schools.

In summary, a potential conflict can be discerned between tradition, preconceived ideas, national syllabi, individual teachers' backgrounds, and their perceptions of the subject. The common denominator seems to be “varying assessment practices,” which is shaped by different factors, implying a potential threat to equity and fairness in assessment.

Comparing CLIL and NAS

Comparing CLIL and the instruction of NAS in Swedish schools reveals similarities and differences between the two teaching contexts. Similarities and differences also appear when comparing Swedish and international contexts. The Swedish application of CLIL differs in many ways from what has been described in other countries where policy documents and CLIL teacher courses are in place, for instance, or where the availability of English in society is much more scarce than is the case in Sweden (Sylvén, 2019). The situation around NAS also differs in comparison to other countries. One reason for this is the heterogeneity of the group and language diversity found among the NAS in Sweden, implying challenges in finding teachers who are proficient in the immigrant students' languages. In other countries, language learners more often represent one or a handful of languages. Below, a summary of CLIL in the Swedish context is provided. Thereafter, the situation for NAS is described, leading to a side-by-side comparative overview in Table 1.10 below.

CLIL

In 2018, the NAE (2018h) conducted an investigation, issued by the government, as to the spread and success of English medium-instruction (EMI) in Swedish upper secondary school. It revealed that schools with EMI or CLIL approaches are most common in or close to bigger cities. CLIL is also most common in college preparatory programs and in subject matter courses in the social and natural sciences. According to the NAE, 1.6% of the students in upper secondary have EMI in at least one subject matter course. The CLIL approaches vary greatly both between and within schools, according to the report, especially as regards how much English is used and teachers' perceptions of the relationship between language and the subject taught. It is noted that from an international perspective, CLIL in Sweden is extremely non-regulated; however, student attainment is above average and grades are generally high already at the outset. Teachers are usually experienced and think teaching in English is rewarding, although at least half of the teachers in the investigation are native Swedish speakers. Teacher certification in the content subjects is lower than general according to the NAE investigation where 64% of the CLIL biology teachers were certified compared with 84% for all teachers and 74% of the CLIL history teachers compared with 87% for all teachers (National Agency for Education, 2018h, p. 91), which may be due to the recruitment of teachers with native English language skills.

One of the biggest research studies on CLIL or EMI in Sweden is called the CLISS project; it was conducted between 2012–2015 (Sylvén, 2019). This longitudinal project compared upper secondary CLIL and non-CLIL students' language learning outcomes in academic registers in Swedish and English, and three different schools' implementation of CLIL. The results show that there is great room for improvement in the CLIL programs at the upper secondary level. Looking at the effects of CLIL on the majority language, Swedish, it was found that the use of academic vocabulary and grammatical metaphors increased the least in the CLIL group in which English was used in all subjects except for language arts. Similarly, there was an increase in linguistic mistakes in the Swedish texts produced by the CLIL students in the same school over the years, whereas the opposite was the case for their non-CLIL peers and for students in one of the schools with a more balanced bilingual CLIL provision. The effects on English proficiency were less prominent than expected. Reading comprehension was the area where a substantial effect could be traced, but other factors, such as extramural exposure to English and the CLIL students' higher proficiency at the outset, influenced the developmental curve for receptive and productive vocabulary so that it actually was the same for both CLIL and non-CLIL students, and no effect could be found due to CLIL.

The CLISS study highlights the need to consider a common policy for CLIL, opportunities for collaboration, scheduling and organization, and teacher competence. As Sylvén (2019) points out, without policy or assessment guidelines it is not a sustainable model to leave the decisions for implementation to the individual teachers. Results from the CLISS project further indicate that multilingual students more often choose CLIL in the first place, and they generally express appreciation of CLIL, as do the CLIL teachers. The CLIL context typically attract high-achieving students, which means that “at present, in Sweden at least, the schools offering CLIL and non-CLIL strands seem to have gone back to streaming students based on their proficiency in English” (Sylvén, 2019, p. 319).

Toth (2018) critically examined language use, policy and practices at a primary EMI school and concluded that English was in a privileged position, implying a native speaker ideal. Swedish was valued as an aid for meaning making, whereas other mother tongues were marginalized. Stakeholder beliefs were found to reflect the same language hierarchy. Toth further found that there were concerns about implications for students’ development of subject-specific Swedish. Swedish was perceived as a potential support in interaction with peers, but the content subjects were taught by native English speakers, and hence English was considered to be acquired for free. Sandberg (2018) investigated teachers’ and students’ attitudes to language use in two CLIL programs at the upper secondary level and notes that whereas new teachers in the natural and social sciences find it challenging to teach in English, more experienced teachers appreciate using English; one reason for this was the wider availability of teaching resources in English. Kontio (2016) and Yoxsimer Paulsrud (2014) found that students in English-medium classes made frequent use of translanguaging strategies. Kontio looked at students in the vocational mechanical program at the upper secondary level and Yoxsimer Paulsrud looked at regular natural and social science programs. Whether or not the translanguaging was seen as an “affordance” (i.e., strategy) varied. Yoxsimer Paulsrud concludes that development of practices and language policy is needed for CLIL (or what she chooses to call EMI), maintaining that in the absence of CLIL policy and pedagogy, it is basically a matter of regular teaching translated into English. Lim Falk (2008), looking at students’ actual language use in Swedish upper secondary CLIL classrooms, found that CLIL was not conducive to learning languages. While CLIL teaching entailed more exposure to English, the students’ active use of English in speech and writing remained very limited. Students and teachers expressed that the use of English as a medium of instruction represented an obstacle. Lim Falk underscores the importance of active language use and interaction in the classroom for the development of a subject-based language, which did not take place in the classrooms under investigation.

In summary, the attitudes, perceptions and uses of the language of instruction vary to a great extent in the studies mentioned above. One shared issue is the lack of policy and stringent pedagogy, which preferably would align with clearly-stated target language goals. These studies represent only a sample, but enough to claim that very little research has been conducted in relation to assessment in Sweden, although one study (Reierstam, 2015; Reierstam & Sylvén, 2019) is summarized in Part 2. A few studies have been reported in the international realm recently (see Part 2), but Massler et al. (2014) call it “a blind spot.”

NAS

No longitudinal research similar to the CLISS project described above has been conducted recently regarding NAS, but a governmental investigation was performed in 2017 with the aim of improving equity in education for these students (SOU 2017:54). The assignment was to investigate how instruction in compulsory school should be organized to meet the needs of late arriving migrant students with a

short history of schooling, since they represent a particularly vulnerable group with regard to academic opportunities. They come at a time when education requires a lot in terms of academic knowledge and literacy, but they have little time to learn the language of instruction. For students who immigrate after the age of 13, the odds of being eligible for upper secondary education are reported to be 70% lower than average. The most challenging subject after Swedish for NAS is English, according to statistics, where only 46% receive a passing grade. According to the investigation, bigger cities have received the largest numbers of NAS, but proportionally very few. Instead, many rural towns welcomed big numbers of NAS after 2015. In certain smaller rural towns, the number of NAS in compulsory school went up by 40% in one year in 2016 (see Part 3 for more information).

The investigation acknowledges the diverse backgrounds and heterogeneity of this group of students. In response to the identified needs of NAS, the investigation suggests organizational solutions to accommodate the varying needs of NAS, including education during school breaks, individualized study plans and access for every student to a study counselor in his or her mother tongue. Today, the organization of education among NAS is reported to be very inequitable, as it depends on access to teachers and school facilities. As regards the national curriculum, it is noted that there are no knowledge requirements and benchmarks below the expected attainment level, although these are needed to track the development and progress of NAS at an early stage. The assessment tools that are already in place need to be further developed, according to the investigation. Moreover, schools' use of the obligatory mapping instrument for NAS literacy and content knowledge (which is currently reported to be unsatisfactorily implemented) needs to be emphasized. NAS often move between towns and schools, and therefore it is important that mapping and uniform communication tools are used to facilitate student progress.

The governmental investigation refers to research and empirically-based instruction but notes that there is an insufficient amount of research to date. It is noted that the distribution of responsibilities as well as collaboration between subject matter teachers, SSL teachers and study counsellors is favorable. Keeping expectations on student performance high and mixing groups of NAS with native speakers to speed up language learning are identified as success factors. Referring to international research, the investigation mentions that teachers need to analyze the disciplinary language in their subjects and have an awareness of language development in the subject. In addition, teachers need to communicate clear goals and use formative assessment. The use of oral presentations and visualizations are also mentioned as good practice with NAS.

More recently, the research focus on NAS has shifted toward concerns about power relations and matters of inclusion and identity (see Part 3). In relation to language, research stresses the importance of communicative dialogic teaching approaches in which the students learn through interaction.

Concerning NAS, students' academic attainment and language development in all subjects have been identified as areas in need for empirically-rooted practices. In contrast to CLIL, a lot of teaching manuals have been presented and the NAE has provided online courses and material with advice on how to teach language in all subjects (National Agency for Education, 2012a, 2012b, 2018e, 2012c); this provides support for students in general, but NAS in particular. A tool called "Build Swedish" (*Bygga Svenska* in Swedish) is available on the NAE's website that provides instructions and assessment material on how to monitor students' progress in Swedish and guide the students toward the next level. Over the last 6 years, new laws have been passed stipulating the assessment of NAS's skills for decisions on placement and to guide the planning of instruction. In relation to language, several studies maintain that the use of the multilingual students' first language and taking previous experiences into account can facilitate their understanding of subject content in the natural sciences (Swedish National Agency for

School Improvement, 2018). Referring to Axelsson et al. (2006), Cummins (2017) notes that the use of “Swedish only” in classrooms solidifies existing power relations. Instead, translanguaging and considering the students’ first language as an asset are promoted. Consequently, whereas CLIL research wishes for more control and stricter policy guidelines, slightly different concerns have been articulated for NAS. Cummins posits that policy, curricula and assessment represent educational structures, which reflect values and ideologies and limit teachers’ choice.

In summary, there is an awareness of the urgent need to provide good education for NAS to improve academic attainment. Many challenges are reported, but several have been addressed with teaching and assessment material, professional development for teachers, and national guidelines that are in line with international research in the field. Still, the decentralized Swedish education system (cf. p. 48f) and challenges in attracting and keeping teachers pose problems, as noted by a report from the European Commission (2019). Issues in reports on the education of NAS concern the need for uniform assessment practices, teachers’ awareness of language in the subjects, and complementary standards or benchmarks for low-achieving NAS during the initial stages.

CLIL and NAS

On the surface, the differences between CLIL and NAS may appear more important than the similarities, but looking at Table 1.10 below, many common features come to the fore. Going back to the integration continuum in Figure 1.10 on p. 39, which ranges from content-driven to language-driven, CLIL and NAS in Sweden would probably be placed in a similar position. Consequently, the instruction of NAS could be labelled as hard CLIL, or they could both be considered partial immersion or EMI and SMI (Swedish as a medium of instruction), respectively. Both are content-driven, and there are no articulated language learning goals. Despite this, Wolff (2009, p. 547) maintains that in the CLIL classroom, language is both content and medium, harmonizing with Swedish curricular goals, which allegedly are very “language dense” (cf. p. 21).

The surge in NAS after 2015 implied a demand for teaching guidelines and consequently has generated new policy documents and recommendations, some concerning language use. For CLIL, there are no official guidelines for upper secondary, but for compulsory school, up to Grade 9, the Ordinance for Compulsory school (SFS 2011:185, Ch.9, §18f) stipulates that only half of the instruction can be in English. Collaboration with language teachers is recommended in literature for both contexts, but in national policy emphasized only for NAS, since again there is no policy for CLIL; instead, different local variants emerge of either bilingual Swedish/English education or English-only approaches.

The subject matter teachers in CLIL are often non-native speakers of English, the language of instruction, whereas the subject content teachers among NAS can be expected to be mostly native Swedish speakers. At least a majority of those teaching the natural and social sciences in both contexts are certified to teach in the subjects, whereas SSL teachers have the lowest proportion of certified teachers, with only 40% (National Agency for Education, 2019d). CLIL teachers often have been the instigators of CLIL programs, and therefore this can be seen as a bottom-up initiative, which is not the case with NAS. CLIL is also based on students choosing to attend programs where English is used as a medium of instruction, unlike the NAS and Swedish. CLIL students often have certain ambitions to seek opportunities in the global arena, whereas NAS seek inclusion in Swedish society. Considering the motives for learning the target language, English versus Swedish, another question arises as regards the role of English in tertiary education (see p. 47). Not only are the NAS at a disadvantage when it comes

to Swedish, they will also be less equipped for advanced university studies due to limited knowledge of English. According to reports from the NAE and a review from the Swedish government (SOU 2017:54a), English is one of the subjects where migrant students have the lowest attainment.

An important factor mentioned for Swedish students' proficiency in English is extramural exposure. Migrant children in general, and NAS in particular, are often segregated both when it comes to housing and introductory programs, leading to little exposure to the language outside of class; this can be described as a difference between CLIL students and NAS. The CLIL students also represent a rather homogenous student group, whereas the opposite is true for NAS, who have varying experiences of previous schooling.

The discrepancies in language education policy between CLIL students and NAS were mentioned above, highlighting the discrepancy in the educational ideologies and motives of translanguaging, additive bilingualism, or multilingualism as plurilingualism. The monolingual motives for CLIL in the Swedish setting imply attaining native-like skills in English, and for NAS the corresponding motive is to become as proficient in Swedish as possible. Translingual ideologies advocate a different agenda, something that needs to be resolved in relation to policy, methodology, language development and management in all subjects and in assessment.

The common features of assessment and instruction in CLIL and among NAS relate to language across the disciplines and academic literacy. The teachers need to know what to assess (i.e., language and/or content) and what language represents in terms of intended learning outcomes in the subjects. Language goals need to be identified and described for various proficiency levels and skills (i.e., reading, listening, writing and speaking) so that the language elements become attainable and able to be communicated, taught, learned and assessed. These matters are discussed further in the synthesis of results from Study I (CLIL) and Study II (NAS) in Part 4.

Table 1.10 Overview of differences and similarities in the Swedish CLIL and NAS contexts

	CLIL	NAS
Language of instruction	English	Swedish
Language ideology	Monolingual/bilingual	Monolingual/multilingual/translingual
Teacher's proficiency in the language of instruction	L2 user	L1 user/Native proficiency
Bottom-up initiative	Yes/No	No
Initiator/Driving force to learn the language of instruction	The student him/herself (choice)	Society
Motivation for the instruction	Preparation for future job market, education, global contacts	Integration, participation in Swedish society, preparation for work and education
Policy for language use	No (yes, Grades 7-9)	Yes (but unclear)
Collaboration with language teacher/mother tongue instructor	No	No/yes
Content-driven instruction	Yes	Yes
Language-driven instruction	No	No/Yes
Students' L1	Swedish (mostly)	Varies
Students' proficiency level in target language, L2	Intermediate/advanced	Beginner
Students' proficiency level in L2 academic language (CC)	Varies	None/Little
Students' proficiency level in L2 disciplinary language (CO)	None/little before taking a course	None/little before taking a course
Students' proficiency level in L1 academic language (CC)	Varies/advanced	Varies
Students' proficiency level in L1 disciplinary language (CO)	None/little before taking a course	Varies
Extramural exposure	Yes	No (yes)

Summative comments on the contextual framework for teachers’ assessment in Swedish multilingual schools

Against the backdrop provided in this first part, a timely short summary is presented below depicting the conceptual and contextual understandings available before we go on to explore the data and findings from Studies I and II. The below figure, which was first introduced in the introduction (Figure 1.3), provides an illustration of the content covered thus far in the thesis (namely, the curriculum and policy aspects, shown in grey text) and the other two parts that the empirical findings will help to illuminate. The national curriculum, course syllabi and other policy documents of different standing (e.g., the Education Act, the Language Act and recommendations from the NAE) provide the basis for teachers’ assessment practices in Swedish education. As regards teachers’ beliefs and actual assessment practices, the data are incomplete and beg the contributions this study aims to provide.

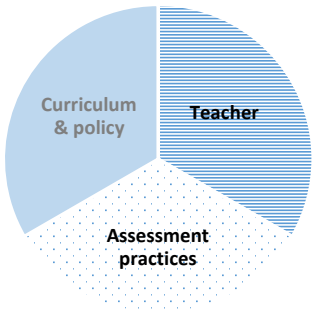


Figure 1.15 Summary of content in Part 1 (curriculum & policy) and areas still to be investigated (teacher beliefs and assessment practices)

To recapitulate, the Swedish education system is largely decentralized, leaving much autonomy to the individual schools and teachers. This has consequences in causing differentiation between individual schools, while schools are becoming increasingly segregated, which has been partly attributed to the freedom of school choice in the system (National Agency for Education, 2018a; OECD, 2015b, 2019a). However, a majority of all Swedish schools—almost 90% of the municipal compulsory schools in 2016 (Hellstén et al., 2020)—have enrolled NAS over the past few years. This explains why teachers in compulsory school request professional development in relation to teaching in multilingual settings (National Agency for Education, 2019f). However, the requests for professional development in assessment have gone down considerably since TALIS 2013 (National Agency for Education, 2019f).

About 16% of Swedish students are not eligible to enroll in upper secondary level education, but for NAS who arrive after the age of 13-15 or during upper secondary, only 29% qualified in 2019 (National Agency for Education, 2019a). The number of students in introductory programs has grown substantially over the past years, including language introduction for NAS. Since 2013, there has been an increase in

NAS enrollment in small rural municipalities (National Agency for Education, 2017b). However, the current teacher shortage makes it difficult to attract teachers (Hellstén et al., 2020), especially to remote areas, and the highest number of uncertified teachers are found in Swedish as a second language courses. Regarding the social and natural sciences subject areas, the number of certified teachers is considerably higher, at about 90% (National Agency for Education, 2018f, 2019d). In policy documents and recommendations from the NAE, translanguaging practices are recommended, advocating the use of the NAS's strongest languages, especially in the beginning. In curriculum policy and a recent statement from the Minister of Education (Ekström, 2020), a monolingual ideology is expressed in order to help students gain proficiency in the Swedish language and thus also increase employability.

Although the special role of English is not undisputed, English is growing in importance, especially in higher education. Swedish youth's proficiency in English as a foreign language (EFL) is among the highest in Europe (Eurostat, 2018) and the extramural exposure and use of English in leisure time is important (Sundqvist & Sylvén, 2016; Sylvén, 2005). This can be contrasted with the statistical evidence regarding NAS's exposure to Swedish, which may be scarce due to segregation in housing areas. The Ordinance for Compulsory School (SFS 2011:185) stipulates a minimum use of 50% Swedish as a medium of instruction to Grade 9 at the compulsory schooling level. At the upper secondary level it is less regulated, and therefore EMI can be used in all courses if the school so chooses (Sylvén, 2019). There is no particular teacher certification for CLIL either and English is used as a medium of instruction by native Swedish-speaking subject matter teachers. For NAS, the Education Act (SFS 2010:800, Ch.3,§12) stipulates the use of study counselors in the students' strongest language to provide the best conditions for academic attainment. However, how this is implemented differs greatly between schools. Increasing collaboration between subject content teachers and SSL teachers has been recommended for NAS, but for CLIL there exist no such recommendations with regard to collaboration between subject matter and EFL teachers.

The syllabi are very language-dense and national high-stakes tests require the use of Swedish, i.e., representing a monolingual policy. The same syllabi, standardized knowledge requirements and national tests are applied to all students, although certain accommodations are accepted. In the natural and social sciences, NAS can do the tests with a mother tongue instructor, bilingual teacher or bilingual dictionary in order to translate certain words. The teacher can also provide help in translating difficult words, but not words that the student must know in order to show relevant content knowledge. Tests or parts of tests cannot be translated and NAS are not allowed to take the national test in English; this is reserved for students who have been in a CLIL program (National Agency for Education, n.d.-a). The mapping of NAS's knowledge in diagnostic tests in literacy, numeracy and various content subjects is stipulated by the Education Act (SFS 2010:800) with the intention of facilitating faster progress for students within content subject tuition before attaining proficiency in the Swedish language. Core content to be assessed and knowledge requirements for how this is to be graded are provided in national syllabi from the NAE. Teacher's "enacted" assessment practices are still underexplored, and questions arise in relation to perceived inconsistencies in other parts of the chain (i.e., language policy and its consequences for individual students, teachers, schools and society).

There are a lack of policy documentation and guidelines for CLIL (Sylvén, 2018), leaving the field open to very diverse and unequal organization. This makes it questionable whether students enrolled in different CLIL schools will have received equal opportunities, and there is no certification to warrant any specific level in the target language. In a similar vein, there is no guarantee for equal treatment for NAS, although this is due to other reasons stemming from decentralization and lack of resources. CLIL students are not at a disadvantage at the outset and attainment is reportedly high (National Agency for

Education, 2018h), but still research shows unsatisfactory effects on academic language (Sylvén, 2019). In the absence of coherent and effective policy and explicit content and language integrated assessment practices for both contexts, the pursuit of good quality practice necessitates additional input from experienced practitioners (i.e., teachers); this is one of the main objectives of the present study.

Part 2 - Assessment in the CLIL context

Study I: Assessing Language or Content?

“I don’t care about the language, that is up to the English language teachers.”
(CLIL biology teacher)

Background

The study “*Assessing Language or Content? A comparative study of the assessment practices in three Swedish upper secondary CLIL schools*” (Reierstam, 2015) investigates biology, history and EFL teachers’ assessment beliefs and practices in three Swedish CLIL schools where English is used as the medium of instruction in subject matter courses. The study was presented in May 2015 as a licentiate thesis. This section presents a summary and abridged version of the study, including some more recent references. For the full original version, see Reierstam, 2015.

Assessment in the CLIL environment has been a blind spot in research for a long time (Massler et al., 2014). Concerns have been raised as regards *what* is being assessed in bilingual instruction, content or language, whether separable at all. In view of what assessment requires of the students, not only in terms of content knowledge but also language repertoire, questions arise as to how students have been prepared to demonstrate their content knowledge. Which language should be used: English or Swedish? Who is responsible for teaching the required language? Or is language learning incidental, a result of language exposure? Assessment conducted in a foreign language may have undesired consequences, as discussed in Part I, since the use of a non-native language may limit the students’ ability to express themselves. Validity in assessment refers to the social consequences of test use (McNamara & Roever, 2006; Messick, 1989), questioning whether or not test results are truly indicative of a certain level of content mastery, and thus how this affects the test taker or student. Conversely, authentic content is often used in foreign language instruction to make language use meaningful. In summary, this leads to questions as to which is being assessed: language or content? Below, the aims and research questions of Study I are presented.

Aims and research questions

The aim of this study is to explore if (and in that case how and on what grounds) the assessment practices differ in two subject content courses (biology and history) in schools where English is used as the medium of instruction instead of Swedish. The focus is to explore if (and if so, how) assessment procedures differ due to teachers’ perceptions of the students’ language proficiency in bilingual education. Another question concerns the English language courses, often called EFL courses: are the course content and thus the assessment tools affected where English is used in subject content courses? A final aim is to look for possible cross-disciplinary similarities that potentially may provide areas of common interest and collaboration in relation to language instruction and assessment across disciplines.

Research questions, Study I:

- CLIL vs. non-CLIL: do the assessment practices differ in the two subject content courses history and biology due to the language of instruction? If they do, *how* do they differ, and on what grounds?
- Are the assessment tools and the course content affected in the English language courses where English is used in subject content courses? If so, *how* are they affected?
- What does the assessment design look like in the different disciplines when it comes to *language, content* and *form*? Are there common features?

In order to find out how language and content are perceived in written assessment practices in the CLIL schools under investigation, a comparative approach was taken. The approach includes subject content teachers (both CLIL teachers using English as a medium of instruction and non-CLIL colleagues using Swedish) together with teachers of English as a foreign language in Grades 10 to 12, as illustrated by the figure below.

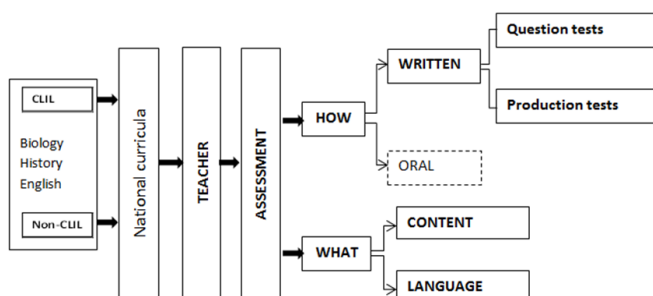


Figure 2.1 Outline of Study I

CLIL in Sweden

The extent and scope of CLIL has been difficult to define, mainly due to a lack of consensus as to what it is. In a recent report from the NAE (National Agency for Education, 2018h, p. 24) it was stated that about 4% of the upper secondary schools have some kind of CLIL provision, ranging from a minimum of one course to an entire program with English as the medium of instruction. English is by far the most common medium of instruction. CLIL is normally implemented at upper secondary schools (Sylvén, 2019), but English is used as the medium of instruction in international independent schools from fourth grade, where a maximum of 50% can be in English, as stipulated by law (SFS 2011:185). At the upper secondary level no such regulations exist, although it has been discussed (National Agency for Education, 2018h). For more on CLIL in Sweden, I refer to Sylvén (2019) and the report from the NAE referred to above. Also, at the end of Part I in this thesis (cf p. 57ff), a comparison was made between the CLIL and NAS contexts in Sweden. In the following, the main focus is on research results as regards CLIL and assessment from Study I (Reierstam, 2015); but first, a summary of previous research in the field is presented.

Previous research on CLIL and assessment

CLIL leans on sociocultural and constructivist learning theory in joining together two complementary views on learning; this, according to Coyle et al. (2010, p. 3), means that “parallels between general learning theories and second language acquisition (SLA) theories have to be harmonized in practice if both content learning and language learning are to be successfully achieved.” Moreover, the emphasis on language pedagogy in research often implies a systemic functional linguistics (SFL) perspective (Nikula et al., 2016), seeing language use as functional to create meaning, thus connecting form and meaning in contexts of use (Halliday, 2014). The question is how the subject disciplinary theories on learning and SLA perspectives are harmonized in assessment in CLIL and related research.

Dalton-Puffer (2007, p. 3) states that “CLIL classrooms are seen as environments which provide opportunities for learning through acquisition rather than through explicit teaching.” According to the European Commission, CLIL can improve the effectiveness of language learning (D. Scott & Beadle, 2014). In a research survey, Nikula (2017) notes a shift in research focus from language learning outcomes to investigating the potential of CLIL to support the development of subject literacies and adopt a truly integrated view on language and content. Previous research on assessment in CLIL comes to the same conclusion (see below), there is a need to distinguish subject-specific language learning outcomes in the content courses. However, Dalton-Puffer (2007, p. 5) notes that despite the word “integrated” in CLIL, there is a “good deal of tension and sometimes conflict between the two areas,” referring to the competition for primacy between language and subject matter courses. Disciplinary strongholds still seem to contribute to a certain positioning, preventing language from becoming an acknowledged learning goal in subject courses and allowing language teachers to free up time from a tight course schedule to collaborate around subject content (see the results section below).

Morgan (2006) advocates new assessment tools for CLIL, since she finds that curriculum criteria and current testing procedures do not accommodate the special language skills acquired by CLIL students. Kiely (2009, p. 4) discusses the purpose of assessment in CLIL as well as the issues of language versus content: “How do we use assessment to manage an appropriate balance in CLIL practice between content and language, such that there is no fear that children [students] achieve less where the learning is in L2?” Others acknowledge the issues in CLIL assessment related to the dual focus on language and content (e.g., Gablasova, 2014; Hönig, 2009; Massler et al., 2014; Wewer, 2014). Gablasova (2014) asks questions in relation to the language used in assessment and lists four options: the students can be tested in the language of instruction (the L2), in their L1, parallel assessment in both languages, or a fourth option implying a mixture of both languages, which could be identified as translanguaging. Problems arise when the students have not learned the disciplinary concepts in their L1, and instruction has not provided them with the academic language skills to express higher-order thinking skills in the language of instruction. No matter which language is used, there will be gaps unless intended language learning goals are made explicit and taught. Otherwise there is a potential threat of construct-irrelevant variance (cf p.45, Part 1) the student is not being offered tools and thus a fair chance to sufficiently demonstrate content knowledge. Gajo (2007) suggests that lists of content-obligatory and content-compatible language should be used, thus promoting collaboration between language and subject matter teachers. In the following, the material and methods are presented, followed by a summary of results from the study in the Swedish CLIL context.

Methods and material

The semi-structured interviews and the questionnaire

The interviews at the first two schools were conducted from March to June 2013. A third school was added, and three more teachers were interviewed from December 2013 to February 2014. The interviews lasted between 16 and 58 minutes, depending on how much time the teachers were able to spend. The paired interviews (one interviewee at a time with the interviewer) took place at the teachers' schools in the teachers' offices or, on one occasion, in the school library (Reierstam, 2015, p. 72ff).

The interview guide included the following themes:

- Teacher background
- Views and experience of CLIL
- Views on teacher's own discipline/subject
- Assessment practices used
- Course and textbook material
- Course outline and plan
- Disciplinary and/or interdisciplinary collaboration

The interviews were conducted in Swedish, except for two where the EFL teachers were native English speakers and it seemed more natural to use English. All the interviews were transcribed, and excerpts were chosen and thematically organized to be used for the analyses. Excerpts used in the essay were translated into English. A questionnaire was used to obtain comparable and corresponding data, as the semi-structured interview format implies certain variation (Reierstam, 2015 p.79, Appendix 2).

The assessment samples

The teachers were requested to present some written assessment samples from one or several of their courses at the interviews, alternatively send them via email after the interviews. Altogether, 42 assessment samples were collected. For the analyses a delimitation was made, to focus on one level in the subject content courses to be able to compare and use the same syllabi. I chose to focus on the first course given in upper secondary school, *History 1b* and *Biology 1*.

For the document analysis, a combination of models was used to describe the features of the tests and the interplay of subject content versus language. Features were selected after consulting literature on test development and assessment in CLIL and the subject disciplines. The language in the assessment items was described in terms of question words and academic function words (e.g., *explain*, *describe*) based on a list made by Dalton-Puffer (2007) and Bloom's taxonomy (Anderson & Krathwohl, 2001). The function words and questions were analyzed in relation to *lower-* and *higher-order thinking skills* and language skills using an adaptation of Cummins' matrix (for more information, see p. 35ff in Part I, where the analytical models are described).

Participants and sampling

In Study I, the participating teachers were found at three upper secondary schools that were all part of the CLISS (Content and Language Integration in Swedish Schools) project (Sylvén, 2019). The three schools, labelled A, B and C in this study, had been invited to be part of the longitudinal research project (2011–2014) and a letter of consent in accordance with ethical guidelines had been signed at the beginning of the project. Consequently, all the participants in Study I were already initiated, and contact information for teachers from relevant subject disciplines was provided by coordinating teachers at the three schools. All 12 teachers, who were first contacted via email, agreed to participate in the initial semi-structured interviews. One teacher retired during the study period, and two teachers chose to withdraw after the interviews. Two other teachers also refrained from handing in assessment samples for the document analyses but remained for the follow-up questionnaire. For an overview of the participants in Study I, see Table 2.1 below.

Table 2.1 Participants in Study I

School A	School B	School C
International school, Swedish curriculum	Two programs (NA & SP)* with one class each per year, mostly in English	One program (SP)* with two classes per year, partly in English
CLIL history teacher	CLIL biology teacher	CLIL history teacher
CLIL English teacher	CLIL biology teacher	Non-CLIL history teacher
CLIL English teacher	Non-CLIL biology teacher	CLIL English teacher
	CLIL/Non-CLIL English teacher	Non-CLIL English teacher
	CLIL/Non-CLIL English teacher	

*NA = The Natural science program, SP = The Social science program

School A is located in a large city and includes students from various language backgrounds, whereas the other two are located in medium-sized towns and have a relatively homogenous student body. School A differs from the other two in that it is an international school where all subjects are taught in English, except for language classes. At school A, English is often heard among students outside of class; in schools B and C, there are parallel non-CLIL classes and the use of English varies both in and out of class (cf. Sylvén, 2019).

Teacher background

The average length of teaching experience for the participating teachers ($N = 12$) was 14 years (three of the teachers did not answer this question). The average CLIL teaching experience was 9.5 years ($n = 9$); one CLIL teacher did not answer this question. In relation to teaching experience, there are two outliers: one CLIL history teacher who had a total of 20 years of teaching experience, but had only been teaching CLIL for 4 years, and one biology CLIL teacher who had a total of 16 years of teaching experience and 7 years of teaching CLIL. Most of the CLIL teachers, including the EFL teachers teaching CLIL classes,

had been teaching CLIL for all or almost all of their professional careers. Besides teaching experience, Table 2.2 below indicates subjects taught, whether the teachers have a degree in English and some supplementary information in the *other* category. The teachers are labelled according to the schools (A, B or C) and were given a random number according to the number of participants at the schools (School A: 1 to 3, School B: 1 to 5, School C: 1 to 4).

Table 2.2 Participating teachers' background

Teacher	Subjects	Year of diploma	Years of experience: teaching/CLIL	Degree in English	Other
A1 CLIL	English, Spanish	2007	6/6	Yes, native speaker	Examiner for ILS, Cambridge exams
A2 CLIL	English, Swedish L2, communications	2005	7/7	Yes, native speaker	
A3 CLIL	History, religion	2007	6/6	No	Teaches geography
B1 CLIL/non-CLIL	English, P.E.	2002	14/14	Yes	
B2 CLIL/non-CLIL	English, Swedish	1982	20/17	Yes	
B3 CLIL	Biology	-	-/-	No	Retired in June 2013
B4 Non-CLIL	Biology, chemistry	1984	28/-	No	
B5 CLIL	Biology, social studies	1996	16/7	Yes	Has taught English
C1 CLIL	English, natural sciences, computers & ICT	-	-/15	Yes	
C2 Non-CLIL	English, Russian	-	15/-	Yes	
C3 CLIL	History, religion	-	20/4	Yes	Master's degree abroad
C4 Non-CLIL	History, social studies	2001	11/-	No	

The section below provides a summary of some of the main findings in Study I.

Findings in Study I

The findings in this abridged summary are presented thematically, merging results from the semi-structured interviews, the questionnaire and the written assessment samples. For a full account of the results, see Reierstam, 2015. For this summary, two themes are used: *language and content in the courses*, and *assessment*. The research questions concern assessment, but teachers' language use and policy in the courses has implications for the validity in assessment in aligning instruction with assessment, and therefore this is covered briefly. Assessment is seen as an integrated part of instruction, as is language. A triangulation of results and final remarks is provided in relation to the three research questions in the concluding discussion.

Language and content in the courses.

In the semi-structured interviews, the subject content teachers were asked how language was dealt with in instruction and assessment. The EFL teachers, on the other hand, were asked if they did any interdisciplinary collaboration integrating subject content from history and biology, for instance, in the English language class, or assisting students with disciplinary language of any kind. One of the EFL teachers at school A raises concerns regarding the effect of English as a medium of instruction:

- (1) **A1/En/C:** We were talking a lot about how we can help our students improve and potentially achieve higher grades, because they're very concerned about grades and rightly so, but it's very hard when you're studying something in another language and you're being, you know, graded on that understanding, that you will achieve the same goals, although the language is not necessarily there in all cases.

However, all the EFL teachers note that they are not "content specialists" and they have a lot of course content to cover, and therefore they have neither the time nor skills to teach or assess other disciplinary language. The teachers at school A mention that they have talked about doing more interdisciplinary collaboration next year. They have started looking at each other's course syllabi, but they mention issues with time constraints. At school A, they have also helped students or colleagues occasionally, but collaboration is generally very rare.

The history teacher at school A notes that he makes clear to the students from the start that it is possible that the use of English as a medium of instruction may inhibit their understanding, which in turn may slow down the learning process. The fact that this very likely means that students will receive a lower grade than they would have if their first language had been used, he sees as only natural. He explains that he sometimes gives students some terminology or sentences in Swedish, especially in their first year of upper secondary school. He feels that they need it and notes that students say they wish more teachers did the same, as not all students in their school are highly proficient speakers of English when they start. The teacher remarks that tricky subject-specific concepts become even harder when English is used as a medium of instruction:

- (2) **A3/Hi/C:** I translate certain words. [...] I do a lot of power points. Sometimes the translation is in parentheses, some words, some terminology and such may be tricky. I have subjects, social studies for instance, where there are lots of words that are tricky even in Swedish to explain and such, so doing it in English doesn't make it any easier.

Disciplinary concepts are often given special attention in the subject content courses, but not the general academic language or the disciplinary genre. Providing students with specific disciplinary concepts is something all subject content teachers claim to do, but how they do it varies.

- (3) **B3/Bi/C:** Sometimes when you know that this is probably a term they don't know, if it is a disciplinary term, then I say it in Swedish too, at the same time, and they get an explanation so it won't be a stumbling block all the time.

The other biology CLIL teacher at school B notes that it is hard for the students to adopt the special vocabulary, but her English colleagues have inspired her to play with words and make word games. She acknowledges that there are a great deal of content-obligatory language and subject-specific disciplinary concepts. She argues that this may be more striking in social studies, her second subject, where the students seem to believe that it is a subject where you can "just talk" using ordinary language, which is not the case. In biology, on the other hand, students seem to expect that there will be many subject-specific concepts. When asked if they deal with general academic content-compatible language in class, the teacher seems a bit startled at first.

- (4) **B5/Bi/C:** I don't know how much I work on that, more than in the classroom where there is a lot of talking. [...] It is based on the key concepts, but I put them in sentences, to create a story, and in that case, it is my way of talking which is either a help or insufficient help. [...] But I have not experienced that this should be a problem.

The teacher says that she provides students with linguistic input when they are listening to her speaking English. She notes that she not only comments on the content, but also on their way of using the terminology in written reports. The CLIL colleague at the same school has a different approach, claiming that the students may just as well use Swedish if they get stuck:

- (5) **B3/Bi/C:** I don't care about the language, that is up to the English language teachers.

One of the history CLIL teachers states that he is very careful not to correct students' grammar in either spoken or written language, since he himself is not a native speaker and makes mistakes. The EFL teachers, on the other hand, have a clear picture of the goal of fostering native English skills, exposing the students to native varieties of English and native use of vocabulary:

- (6) **A1/En/C:** That's, you know, the real test, if they're producing them afterwards. Modelling, I think, good behavior, or good language behavior I guess in a sense; what kind of words would I use, or a native speaker, surrounding a certain text type?

One of the non-CLIL English teachers actually mentions that he likes to choose topics that align with the students' program profile in order to get them ready for university studies, he claims. Hence, he reports collaborating with a history subject content colleague, reading and discussing *Animal Farm* by George Orwell in English in the English language class when the students were working on the Russian Revolution in history class. The teacher states that the students like the use of authentic content in the EFL class. As regards the teachers' perception of the students' proficiency level in English, the EFL teachers at the international school note a slightly different view compared with the language teachers at the other two schools. Whereas the latter claim that their students are very proficient English language users, the teachers at the international school, who themselves are native English speakers, mention that they use the first weeks of upper secondary to help the students work on their spoken interaction, since

this will be needed in all courses. As noted above, one of the history CLIL teachers also noted that the students are not as proficient language users in the subject content course.

Regarding the language used in class, the situation in the subject content courses varies as reported by the teachers. One of the biology CLIL teachers declares that all the communication on her part is in English and that the students try, but sometimes use Swedish if they feel unable to do it in English:

- (7) **B3/Bi/C:** It is better that they say something, even if it is in Swenglish² or Swedish, than them not saying anything. It can be a bit tricky, but it is not language which is the main thing, it is still biology that is the main thing so to speak.

She explains that there have not been any requirements from the school management to speak only English, even though the school has an international profile, but she argues that since the language may constitute a hindrance, the students should get all the help they need. She acknowledges that when the content theme has been particularly difficult, such as when they were working on the anatomy of the human body, some students asked for and received material in Swedish.

Both history CLIL teachers (schools A and C) claim to use only English in the classroom, even though Swedish may be used occasionally in individual conversations with students about their achievements.

- (8) **C3/Hi/C:** When I switch into Swedish it is mainly for student feedback on grades and course evaluation so there are no misunderstandings—if Swedish is the first language of the student, that is [...] If the students want to ask something during class, and want to ask in Swedish, I accept that of course, and reply in English. It has to do with not inhibiting them from asking questions.

Thus, the CLIL subject matter teachers express different views regarding the use and role of language as part of course content and assessment, as do the English language teachers.

Assessment

When it comes to assessment, teachers note that it is not only a matter of the students' receptive skills and understanding. Several teachers acknowledge that they have been concerned about the students' ability to show content knowledge using English, as this history CLIL teacher:

- (9) **A3/Hi/C:** Well, it is a bit tricky, because they are supposed to show their understanding, that's the thought, and knowledge of certain concepts and so forth, and sometimes when their writing is confusing, and it is not correct, it makes you wonder if it is a matter of issues with the language or problems with understanding.

The teacher's solution to the problem, similar to that of one of the other CLIL teachers, is to approach the student and ask. He describes it as a written dialogue, where he comments in the margin if there is something he does not understand in the student's text. Then, he may ask the student for clarification and make new comments before handing it back to the student when he has some more information; this process may take a couple of weeks.

² Swenglish is a term coined to denote a mix of Swedish and English

When asked if the use of appropriate vocabulary is part of the assessment, the other history CLIL teacher comments that he finds no room for that in the knowledge requirements, but he brings it up in his feedback to the students.

- (10) **C3/Hi/C:** If I am unsure [what a student means] I start a discussion: “I understand what you mean, but can you think of this next time?” But English proficiency cannot be part of an examination just as Swedish should not be part of an assessment in history.

One of the EFL teachers maintains the same thing: that assessment in biology, for instance, is about assessing knowledge of the subject, not the language. One of the non-CLIL teachers, however, states that assessment in history is based on how articulate students are and how proficient they are in storytelling. He acknowledges that essays are demanding to grade, and therefore many teachers prefer written question tests, which are easier. He uses a variety of tests and essays. However, he believes it is difficult to show evidence of different proficiency levels in a regular question test, since there is not enough time to use analytical skills. The teachers refer to what is required in syllabi when choosing assessment format and the character of the subject. The history teachers, CLIL and non-CLIL, mention the narrative character of history, and therefore storytelling is mentioned as appropriate as a disciplinary genre together with the more general argumentative academic genre that will be required in higher education. The non-CLIL history teacher notes that, unfortunately, there is no time to work on the writing process, since the subject syllabus includes so much content that needs to be covered.

The EFL teachers generally report working on the writing process using a range of genres, including argumentative essays. One of the English CLIL teachers reports that he looks at both language and structure and content. One of the other EFL CLIL teachers mentions the use of a portfolio to gather a variety of texts, but claims to only look at language, not content, as in fluency and coherence. However, he notes that with 130 students it is not possible to provide individual feedback, since it would be too time consuming. All EFL teachers note that they are already working with essays and academic writing during the first year at the upper secondary level, but that it becomes more and more prevalent in the advanced classes in Grades 11 and 12. One CLIL English teacher mentions that it is a matter of finding relevant and appropriate topics for that level, but there is no collaboration with subject matter colleagues to integrate authentic themes from the content courses.

One of the problems the teachers note with written assignments and essays is the possibility of plagiarism, but one of the CLIL teachers claims that the language the student has used reveals if it is the student's own words or not. Oral exams could be an alternative, but he mentions other issues, such as how one can evaluate what a student knows if he or she does not dare to speak in oral assessment, or in the case of group work, how to know who has contributed and with what. These are some of the reasons he believes written assessment is easier. He states that students prefer written assessment over oral as well. When questioned why, he mentions that the language barrier might be one explanation, but also that students are used to written tests from previous schooling. One of the biology CLIL teachers states that she allows the CLIL students more time on tasks in assessment situations, which is an accommodation due to the use of English, according to her.

All of the teachers, both CLIL and non-CLIL, favor written tests. One of the biology CLIL teachers explains why she likes written tests, but also lab reports:

- (11) **B5/Bi/C:** I think written tests have the advantage that you [as a student] have the possibility to sit on your own and to really express yourself, to use your language. Therefore, I think tests at the end of almost all themes are good. However, labs are also opportunities for assessment, [to show] their [the students'] way of applying a method, and that is in the classroom, which they have to receive feedback on as well.

One advantage with written tests is noted by the teachers: students can use their language. The biology CLIL teacher devotes time after every test for oral feedback, all of which is done in English. She stresses the importance of feedback for students' development but adds that it also teaches her something about the way students think and, most importantly, about their goals. Her CLIL colleague remarks that she tries to have some oral presentations as well and tries to find time to talk to students in small groups, since personal communication is important for assessment purposes. She finds it hard because there is not enough time, and some students are not very talkative; the language might be a barrier, even though they are allowed to use Swedish.

When discussing the possible hampering effects the English language may have on students' ability to present their knowledge, one of the CLIL teachers says that since the students have received all of their instruction, practice and teaching materials in English, they should be able to do it in English.

- (12) **B5/Bi/C:** The question is whether the instruction has meant that the teacher has transmitted his or her knowledge but not put the students in a position where they have practiced how to present their knowledge [...] I believe that students sometimes find it difficult to present their knowledge regardless of language.

The teacher clearly makes a distinction between teacher transmission of knowledge versus engaging students to use the knowledge but does not explain how meaning is created in use other than focusing on putting key concepts into sentences (cf. excerpt 4 above). Both biology CLIL teachers mention progression in students' knowledge when constructing test items and when assessing; as the highest grade, A requires more elaborate answers or higher-order thinking skills:

- (13) **B3/Bi/C:** Well, you try to make the test items graded by difficulty, according to the grading system that is [...] so you try to make some A-questions, some C-questions and some E-questions, maybe, and then it shows how well they are able to solve those. [...] Sometimes you do it with points as well, but that maybe shows the same results, which is as good.

In order to answer the more demanding questions, what does this CLIL teacher believe the students need to know? The teacher is open to "translanguaging," which is not a problem in this case since she is bilingual in both languages in question.

- (14) **B3/Bi/C:** I don't require that they must know specific concepts when it comes to the kidney; renal pelvis for example. They probably know "kidney", but if they don't remember renal pelvis, they can write it in Swedish. [...] I don't require that they have to know specific words, that is not the main point, they need to know relationships; how does the kidney function—describe [...] If they don't know all the words in English, then it works just as well in Swedish. They need to know how to describe what happens.

A contrast can be found between the focus on knowledge as concepts previously, whereas here in Excerpt 14 it is mainly about procedural knowledge as in verbal skills, i.e. to describe.

Below are three sample items from biology. The first item has a selected response (SR) format and requires matching; this question generates an E grade or a maximum of a C. The second example has a constructed response (CR) format where language production is required; this question can generate all three grade levels: E, C or A. Excerpt 1 and 2 are from biology CLIL. The second item contains two of the most frequently used function words: *describe* and *explain*. The question words *why* and *how* require more language as well as higher-order thinking skills. The third example, Excerpt 3, is from a non-CLIL test and is also an essay-type CR item, which may require a lengthier answer of up to one page, generating an E, C or A grade. This could indicate a slight difference between the CLIL (examples 1 and 2) and non-CLIL (example 3) questions. Whereas most questions are identical, the biology CLIL tests do not contain the lengthier essay questions, although lab reports, which require language production, are used in both.

Assessment sample 1: Bi CLIL:

Combine pictures with the right characteristics. Put the right letters in the boxes underneath the phyla below. E/C (5/2)

A) No tissue

C) Coelomate

E) Diploblastic

G) Asymmetrical

I) Sessile




B) Tissues

D) Acoelomate

F) Radially symmetrical

H) Bilaterally symmetrical

J) Triploblastic

Characteristics phylum 1	Characteristics phylum 2	Characteristics phylum 3

Assessment sample 2: Bi CLIL:

The nitrogen cycle (E, C, A) 2/2/4

- a) **Why** is it so important for a living organism to be part of the nitrogen cycle?
What is N_2 used for in life?
- b) Nitrogen, N_2 , is a major part of the atmosphere. **Describe and explain** how nitrogen can transform into forms for living organisms to use.
- c) **How come** the level of N_2 stays the same?

Assessment sample 3, Bi non-CLIL:

Evolution (E/C/A)

- a) **Give an account of how** scientists think life evolved during the early years of earth's history.
- b) **How** has it been possible to illustrate this in an experiment?

Whereas the biology tests contain some questions that are context-embedded and provide visual scaffolding, as in the first example above, the test items in two of the history tests are more similar to the items in samples 2 and 3 (i.e., they are CR-type questions without any multi-modal support). One of the three history teachers uses many pictures in one of the tests, but not in the rest. The biology teachers work at the same school (B) and mention that they collaborate, and therefore the CLIL tests appear to be translated from Swedish to English to a large extent. Two of the history teachers also work at the same school (C) but have very different assessment designs. It seems to have more to do with personal preferences than the CLIL versus non-CLIL approach, although one of the CLIL teachers offers some cues in the first test, providing the initial letter of the desired responses, which might be perceived as some kind of accommodation (see example 4 below). The non-CLIL teacher at the same school offers some context-embeddedness in a couple of the initial test items in a test as well, using an SR-format question (example 5 below) that does not require any language production (I have translated it to English as the initial item was in Swedish). This item only generates an E, whereas the level generated by the points awarded for the CLIL item, or the total scores at the end of the test, is not as transparent.

Assessment sample 4, Hi non-CLIL:

Place the following events **in chronological order**. The timeline starts year 1000 B.C. and ends 1500 A.C. (E)

- a) The Black Death
- b) Christianity becomes the official religion in the Roman Empire
- c) Alexander the Great conquers the Middle East
- d) Sweden, Denmark and Norway are united in the Kalmar Union
- e) High Middle Ages start in Europe
- f) Julius Caesar becomes a dictator in Rome
- g) Athens is the leading city-state in Greece

Assessment sample 5, Hi CLIL:

What was the popular name for the young Macedonian king who was in charge of the Macedonian troops and personalized this development? He had also a very famous teacher, almost as famous as himself. **What** was his name? (4)

Answer:

The king: A _____

His famous teacher: Ar _____

Example 6 below is from one of the CLIL tests. It is found in a test on World War II and is context-embedded but in turn contains several sub-questions. The answer is supposed to be analytical and argumentative. The test item requires both higher-order thinking skills and more language skills, since a thorough answer containing both subject-specific concepts and content-compatible academic language is needed. How example 6 is graded and assessed is not made clear, but the teacher claims to use rubrics and therefore the item and the whole test are probably assessed holistically using all three levels, E, C and A.

Assessment sample 6, Hi CLIL:

Level requiring a more **analytical** answer. Try to answer as thoroughly as possible and, if needed, point out what your opinions are.

The Holocaust is the name given to the atrocities during which the Nazi regime in Germany systematically killed millions of people, not least Jews. It is easy to see the responsibility of the Nazis in this. But **what** responsibility did the Germans as a group and as individuals have concerning the Holocaust? And **what** about other countries... to what extent could we say that other countries had a responsibility and could have acted in ways that perhaps could have prevented or stopped the Holocaust?

Try to **give arguments** based on for example...

- your knowledge of people's awareness of the treatment of Jews
- your knowledge of how other countries acted towards Germany and Hitler

As expressed above, the focus of one of the biology CLIL teachers is on relationships and procedures, not concepts in English; the students can use Swedish if they like. However, the non-CLIL colleague as well as the other CLIL teacher state that the students need to know the meaning of concepts in Swedish, but also provide explanations and compare. Below is an example from the non-CLIL teacher, describing the intended learning outcomes in biology and the structure of the exams:

- (15) **B4/Bi/nC:** There is a lot of problem solving, and you build on cases explaining phenomena which have occurred, comparing different systems. And then there are warm-up questions, if you think of tests, where you have to know the meaning of concepts, there are many concepts and models.

One of the history teachers mentions that the new national grading criteria from 2011 require more analytical skills even at the most basic level in order to receive the lowest passing grade, an E.

- (16) **A3/Hi/C:** We try to structure it, [...] the essay, so there is, well it depends a little bit, an introduction, some sort of descriptive body, an analytical part and a conclusion, and then there is a bibliography. I introduce the different parts, and here in the analytical part [...] it often deals with a certain type of knowledge, or levels of progression.

The teacher explains that he uses rubrics. The other CLIL teacher declares that he dislikes matrices and manuals and sees a problem in what he feels is mechanization and bureaucracy in assessment. He finds the dialogue with the individual student to be the most important tool and feels it works just as well without rubrics. The national course goals are enough for feedback, according to him. When asked how to make progression visible and how to explain different proficiency levels to students, he states that this becomes visible in the dialogue during the development of a text, where peer feedback as well as teacher feedback help develop the relevant skills. How well students handle the questions that appear while working with a text also constitutes a variable for grading, according to the teacher.

The questions in the assessment samples were analyzed and a frequency count was performed of the academic function words and question words used in the test items. See the two tables below for a comparison between written assessment samples in CLIL biology versus non-CLIL biology and CLIL history versus non-CLIL history. The function words are organized in a tentative order of cognitive and linguistic demands, with the easier at the top and the more advanced toward the bottom, although boundaries are fluid, as noted in Part I.

Table 2.3 Function words used in the test items

Function words	CLIL biology	Non-CLIL biology	CLIL history	Non-CLIL history
Encircle	1			
Match		1		
Put in order				1
Write in the right place	1			
Name	1	2	4	
State	6	6		
Mention			5	
Mark	1	1		
Give an example	1	1	1	
Define	1	2		1
Describe	2	1	4	1
Discuss	1	1		2
Explain	12	14	2	4
Draw		2		
Show	4	2		1
Compare	2	1		1
Motivate	1	2		
Analyze		1		1
Give arguments			1	

Table 2.4 Question words used in the test items

Question words	CLIL biology	Non-CLIL biology	CLIL history	Non-CLIL history
What	7	3	43	2
When	2	1		
Where			1	
Who			2	
Which	3	5	1	2
How	6	3	2	1
Why	2			1
Yes/No question	1		2	

The purpose of the frequency count is to provide an overview for a comparison between subjects and contexts (i.e., CLIL/non-CLIL). However, the sample is small, and therefore generalizations cannot be made. As shown by the tables, there is no clear difference between CLIL and non-CLIL as regards the use of the less demanding function words. Similarly, the function word *explain* is the most common in both biology contexts, but not as common in history. As seen in the assessment sample above, an analytical answer was expected in one of the history CLIL samples, but the word “analyze” was not used in the question; instead, students were asked to give an “analytical answer.” It was also clear that

the question word “what” can require both simple one-word answers as well as lengthier, constructed responses.

In the questionnaire (N=9), the teachers were asked what they thought was most important when assessing students’ knowledge and skills in their subjects. The results are summarized in Table 2.5 below. The teachers could choose as many alternatives as they wanted, but again, the question prompted them to make an evaluation in relation to importance. The English language teachers are presented at the top, followed by the biology teachers and finally the history teachers. The teachers are listed according to their assigned numbers except for B4 and B5, who have been switched in order to start with the CLIL teacher in both subject disciplines. (Teacher B3, Bi CLIL, had retired by the time of the questionnaire.)

Table 2.5 *What teachers claim to be most important when assessing students' skills*

Teacher:	Written & oral performance	Subject content	Use of subject-specific language (CO)	Use of general academic language (CC)	Mastery of various forms of expression and IT	Linguistic accuracy	Linguistic complexity	Mastery of disciplinary written genre	Analytical skills
	A	B	C	D	E	F	G	H	I
A1 CLIL English	X				X				
A2 CLIL English	X								
B1 CLIL/non-CLIL English	X		X	X	X	X	X	X	
B2 CLIL/non-CLIL English	X			X			X	X	
B5 CLIL biology		X	X				X	X	X
B4 non-CLIL biology	X	X	X					X	X
A3 CLIL history		X	X						X
C3 CLIL history	X	X						X	X
C4 non-CLIL history	X	X	X						X

Many aspects could be discussed in relation to the character of the subject and teachers’ individual preferences, but as noted above, the sample is small. In Table 2.5 above, the feature D, *use of general academic language* has been highlighted, since this is downplayed by all teachers except for two of the

EFL teachers. The result is expected given the nature of the question and the more obvious importance of other features in the subject matter disciplines. However, as noted in Part 1 of this thesis, for reasons of fairness in assessment in a CLIL context, it is important that the students are given an opportunity to learn general academic language. The two EFL teachers' responses may indicate that they are targeting language in their classes that may be helpful in expressing content meaning in the subject matter courses. The character of the general academic language used and taught in the EFL classes would need to be investigated further. A relevant question is thus how analytical skills, which are claimed to be important by all subject content teachers, are identified in terms of general academic language features, and if these are indeed taught in the EFL classes for students to use in the subject content assignments. The different alternatives listed in Table 2.5 all include language features to a varying degree. They bring to the fore the relationship between the notions of language and content, which are given different weight and are interpreted differently depending on subject discipline. Coyle et al. (2010) note that content covers desired learning outcomes in relation to knowledge, skills and understanding, which means that language is an integrated part of content. Ball et al. (2016) make a distinction between concepts, procedures (including skills) and language in skills, arguing that concepts can be part of content (Ball, n.d.): "The content (conceptual) is learned through a procedure (skills) which requires a certain type of framework (language)." The so called "framework" would thus include general academic language. For more, see Reierstam, 2015. Below, the findings are discussed in relation to the research questions.

Discussion

CLIL vs. non-CLIL: do the assessment practices differ in the two subject content courses history and biology due to the language of instruction? If they do, how do they differ, and on what grounds?

When a non-native language is used for the assessment of subject content in schools, it is important to make sure assessment is fair to students who are not only learning the subject, but also the language in which it is taught. As stated by Llinares et al.(2012), it is important for CLIL teachers to be clear about the possible effect of language on students' performance when assessing subject content. Therefore, the aim of the first research question is to investigate whether (and if so, how) teachers' assessment practices are affected by the language of instruction in the CLIL context.

Teachers' perceptions of language appear somewhat ambiguous. Subject content teachers as well as English language teachers profess that it is not up to the content teachers to assess language. Some subject matter teachers in the CLIL provision claim that they do not require the students to know subject-specific concepts in English; instead, they allow them to use Swedish or mix English and Swedish as long as they make themselves understood, which might be identified as a translanguaging practice and a kind of accommodation. What matters is whether the students can describe relationships in biology, and assessment should provide opportunities to mirror students' progression, as shown in more advanced analytical competence and language. Progression in subject knowledge and skills hinges on problem solving, the ability to build cases, form hypotheses, draw conclusions, explain and compare different systems, according to the biology teachers. CLIL or not, the biology teachers refer to the knowledge requirements expressed in the subject syllabus. One of the CLIL history teachers mentions that the purpose is for students to show their level of understanding, which is difficult when English is used. A common solution in history seems to be not to alter the construct or the assessment format or administration by providing test accommodations, but rather to approach the students after the test has been graded to ask for clarifications and make adjustments in the test scores based on the interaction with the student. However, CLIL or not, one of the teachers notes that students seem to find it difficult to present their knowledge, regardless of the language.

Looking at the assessment samples, differences in assessment format or use of different function words seem to have more to do with the character of the subject or the preferences of the individual teachers than the language of instruction. Biology teachers favor question tests with a variety of SR and CR items. The history teachers favor the production of lengthier texts in order for the students to use the disciplinary language and show either argumentative analytical skills or narrative skills; history, according to one of the teachers, is a narrative genre. Some test accommodations can be noted, as the teachers mention extended time on tests and that the students are allowed to use Swedish at times, both when speaking in class and choosing texts to read, and when responding to questions on a test. In relation to the language of instruction, the CLIL teachers also mention that they need strategies to make oral follow-ups and ask students for clarification, since it is unclear at times if the students have understood or have insufficient language when taking a test in English. Oral assignments are mentioned as a strategy, but all teachers claim they prefer written tests and assignments, as do the students, according to the teachers.

The CLIL profile does not seem to have much effect on either how assessment is done or on what is assessed, as claimed by the teachers. Although they are very certain that language is not part of assessment, they maintain that assignments where the students get to use language are needed to assess certain skills. The teachers differ in how strict they are as regards the use of English. Two CLIL teachers (one biology and one history) express that English should be used, whereas the other two CLIL teachers

are more open to translanguaging or the students' mixing of languages. In the questionnaire, one of the CLIL content teachers (in biology) stated that linguistic complexity was important in assessment. Both the CLIL and non-CLIL biology teachers reported that mastery of a disciplinary genre is important, whereas only one of the history CLIL teachers shared this opinion. All content teachers stated that analytical skills are important in assessment tasks. None of the subject content teachers (CLIL or non-CLIL) mentioned the importance of general academic language or mastery of various forms of expression, such as ICT.

Are the assessment tools and the course content affected in the English language courses where English is used in subject content courses? If so, how are they affected?

The EFL teachers in the CLIL context put a lot of emphasis on the national tests, but also mention varying assessment formats: grammar and vocabulary quizzes, oral assignments, book reports and different written genres such as narrative, exploratory and argumentative texts, which work as long as the topic is relevant to the student. One of the non-CLIL EFL teachers mentions integrating subject content from other disciplines to prepare for university courses. At the international school (A), one of the teachers comments that sending written drafts back and forth to revise and improve is too time consuming with 130 students. He thus suggests that this could be an area of future interdisciplinary collaboration, but he quickly adds that he is not capable of assessing content. The teachers in school A mention that they have helped students when choosing a topic for comparative essays in subject content courses. They also emphasize oral interaction at the beginning of the school year to help prepare the students to interact in their other courses. At school B, the biology teachers note that they have received some input from EFL colleagues concerning vocabulary activities that can be useful when working on subject-specific concepts. All EFL teachers report that they have a pretty full schedule and usually follow a pre-made plan for what needs to be covered, including preparations for national tests at the end of the year. There is no time for interdisciplinary collaboration or integration of content from other courses. The EFL teachers also express that language should not be assessed in the content courses, just as subject content should not be assessed in theirs.

What does the assessment design look like in the different disciplines when it comes to language, content and format? Are there common features?

One common feature is the written genre; however, subject content teachers as well as EFL teachers are concerned about plagiarism if they let students do assignments at home. Producing academic text is identified as an assessment format suitable for collegial collaboration and subject and language integration. The use of authentic and relevant topics and various skills (e.g., critical thinking) to analyze and compare is a common feature. Although none of the subject content teachers mentioned the importance of general academic language when assessing students' skills, two of the EFL teachers did, which could suggest grounds for collaboration across disciplines. Another common feature is that only one teacher, an EFL teacher, expresses the importance of linguistic accuracy, thus pointing to a communicative functional language view in all subjects rather than a focus on form. All teachers, regardless of discipline, share the weight attached to covering course content as expressed in syllabi and the knowledge requirements, indicating a shared accountability culture.

Concluding remarks

To conclude, the CLIL (Content and Language Integrated) approach as seen in this study from the Swedish context is content-driven rather than language-driven (see integration continuum, p. 49). The target language, according to the CLIL teachers, is English, hence implying a monolingual rather than bilingual pedagogy in the present study. The focus in the content courses is on teaching subject concepts in English rather than in both English and Swedish. If and when Swedish is used, it is as an accommodation rather than being taught consistently, as reported by the teachers. The national curriculum states that academic language skills should be fostered in Swedish, but advanced English skills are part of the curricular goals as well, revealing the status the English language holds in Sweden. The teachers note that the CLIL students are motivated and academically driven and the EFL teachers express that the students are very proficient in English. The content teachers, however, state that it is hard for the CLIL students to adopt the special subject vocabulary, that the use of English slows down the learning process, and that the students sometimes are hesitant to speak and prefer to write, and most likely will achieve a lower grade in the content course. Several of the subject content teachers maintain that they do not correct language since they are very conscious of their own deficient non-native language skills. The native EFL teachers, on the other hand, express that they need to model native language usage, since the students' academic language is not always proficient. The EFL teachers are the only teachers who report that general academic language is important in assessment, although the subject content teachers are concerned about how language may interfere with student attainment in their courses, as exemplified previously.

The question that remains unresolved is whether it is up to the content teachers to foster subject literacy skills, and if so, how? As an integrated part of engaging with subject content, or in collaboration with EFL colleagues? In the absence of policy guidelines regarding language use and explicit language learning outcomes, validity in assessment can be questioned. This is especially the case regarding the underlying threat of construct-irrelevant variance, represented by language- and subject-specific genre that are not taught, but assessed. Incidental language learning implies a risk of unequal access to subject content and to inequitable grades. When subject content teachers maintain that they do not assess language, it depends on what they perceive of as language in relation to subject content, i.e., concepts, procedures, skills or general academic language (cf. Ball et al., 2016). After reviewing the assessment samples presented in this research, it is unfair to say that subject knowledge can be displayed without language; instead, there are many language features, levels, genres and registers involved in creating meaning. If language is part of the assessment such as intended target language use, whether it be in lab-reports or argumentative essays, the related language needs to be taught to ensure appropriate and fair assessment. Additionally, it needs to be established by whom it will be taught: either the EFL teacher in the language arts class, or the subject content teacher, or both in collaboration.

Part 3 - Assessment in the migrant context

Study II: Assessment in Multilingual Schools

“They can ask if they want to know what I am after in the question; sometimes the question may be unclear. It is not easy to formulate questions that give equal opportunities to all students”. (NAS teacher)

Background

This chapter serves as an introduction to Study 2, providing a background to the current Swedish situation pointing to the relevance and need for an investigation on the topic of assessment among NAS. Research in the field is particularly important since questions of equitable assessment and school segregation have come to the fore lately. In the following, some data are first provided to describe the context (i.e., how NAS have been welcomed and challenges in relation to academic achievement and inequity in assessment), before introducing the research questions and aims. Given the specific challenges in this educational setting, a literature review follows, providing some examples from Swedish policy documents, investigations and research.

In 2015, Sweden experienced an unprecedented increase in the migrant student population. A few years before, the number of immigrants had grown substantially. The number of NAS in compulsory school doubled between 2012 and 2016, but decelerated after 2017 due to a more restrictive migration policy (Mörtlund, 2020). Today, at the time of this study, every fourth student in Swedish compulsory schools has a foreign background, and in upper secondary the figure is 30%, according to NAE statistics (2019b, 2019c). However, by 2014 the OECD had already noted that Sweden’s migrant population had been growing for many decades, putting Sweden among the OECD countries with the largest foreign-born populations. The OECD (2014) concluded that integration of immigrants and their children is therefore of key importance, not only for the Swedish economy, but for society as a whole. The governmental investigation reached the same conclusion (SOU 2017:54).

By 2009, before the big surge in NAS, the Swedish school inspectorate alerted that the ability of the schools and teachers to meet the needs of language learners varied a lot, which raised concerns as regards inequality in the system (Swedish National Agency for School Improvement, 2007; Swedish School Inspectorate, 2009). This is particularly troublesome given the “compensatory mission” explicitly stated in the Education Act, in which it says that education must take individual needs into account and strive to level out differences in the students’ conditions (National Agency for Education, 2016). In Sweden, all schools report their students’ final grades in Grade 9 to the NAE, and in 2016, after the big surge of NAS, there were indications that schools had handled assessment and grading very differently. The following year, the differences were not as prominent (National Agency for Education, 2018b, p. 7). This can be explained by an absence of guidelines in the school ordinances on how to report grades at the end of Grade 9 if there is no formal decision on an adjusted schedule, according to the NAE, since the grading regulations assume that students have been part of mainstream education and are thus to be

graded in all subjects. Cummins (2017) notes that the Swedish decentralized and partly privatized education system complicates the organization, coordination and implementation of school reform and policy, thereby making the newly-arrived situation more difficult.

Students with a migrant background form a very heterogeneous group and the variation is big when it comes to country of origin, language, socioeconomic background and schooling, all of which have implications for academic success (Bergendorff, 2014; Bunar, 2010; National Agency for Education, 2011). Bergendorff (2014) notes that unlike countries such as Canada and the U.S., which have more homogenous migrant groups, the newly-arrived in Sweden are much more heterogeneous; Swedish teachers therefore cannot possibly be expected to have knowledge about all cultures and variations in student background. Bigger differences are displayed between the NAS in Sweden due to background than the OECD average (Cummins, 2017). In 2015, Swedish schools had NAS from 165 different countries, with the biggest groups represented by asylum seekers from Syria, Somalia and Afghanistan. Sometimes, as many as 54 different mother tongues can be found among the NAS at the same school (SOU 2017:54).

Time of arrival and age represent important factors in relation to students' academic progress. The average age of many asylum seekers has increased (Fratzke, 2017). Migration before or after original school entry has been reported to correlate with school attainment. The poor results of students arriving at an older age are well documented (Cederberg, 2012; National Agency for Education, 2005). Reports show discouraging numbers reflecting very low academic attainment among students with a migrant background. Of the students who arrived in Grades 1 to 5, 67% became eligible to apply to upper secondary school in 2018, but among the newly-arrived who came in Grades 6 to 9, only 28% qualified; in 2016, the same figure was 16% (National Agency for Education, 2019e). In 2016, the government issued an investigation to ensure that more NAS qualify for upper secondary education; some of the resulting recommendations are presented in the following. Compared to many other European countries with many migrants and asylum seekers, Sweden has had a political agenda that explicitly seeks to give immigrants as many opportunities for education as possible. However, researchers have been critical regarding the gap between well-intentioned measures and inefficient implementation (Cummins, 2017, p. 287).

Some of the measures that have been suggested include the initial use of the student's strongest language (sometimes referred to as translanguaging) and diagnostic assessment (the so-called "mapping" of the students' literacy and subject knowledge to be able to plan instruction). It has also been suggested that schools should focus more on the students' strengths rather than their deficiencies. The school inspectorate (2009) has criticized schools for focusing too much on students' learning of Swedish instead of providing study counselors in their mother tongues to help them progress in subject matter courses. (SOU 2017:54). In a country note with regard to the results of PISA (Programme for International Student Assessment), the OECD (2015c) expressed that the performance gap in Sweden between immigrant and non-immigrant students in science is larger than average. The language skills needed in the higher grades require a lot from the students; this is a battle they often cannot win, according to Hajer and Meestringa (2014), resulting in low self-esteem unless the subject matter teacher knows how to help the students. As regards language policy and a demonstrated unequal access to study counselors, Kaya (2017) notes that for NAS who speak a language like Arabic, there may be access to a mother tongue instructor or interpreter in the assessment situation. If, however, the students speak a minor language, they may not get the same opportunity to use their strongest language, thus leading to inequity in assessment and grading.

Even though the situation became exceptionally challenging after 2015, Sweden's experience of language learners and migrant students in education goes back many years. Immigration began in the 1950s with labor migration and continued into the 1980s, sustained by a liberal refugee policy. Although a lot has been written on migrant students in education over the years (National Agency for Education, 2005, 2011), circumstances were new. Much attention was given to the fact that the NAS were spread unevenly across the nation, with some schools and municipalities welcoming large numbers and others, none. Some schools, especially in larger cities, had experience with language learners; for others, it was a fairly new phenomenon. Certain schools suddenly found themselves with classes where more than 60% of the students were newcomers. The proportion of students with a migrant background sometimes differs between schools within the same municipality, with 95% in some schools and 4% in others (Dagens samhälle, 2017). In 2016, 40% of the NAS ended up in 10% of the municipalities (SOU 2017:54, p. 55). In 2015, schools and municipalities were expected to put an organization in place while at the same time arrange instruction for the arriving students. As noted in the investigation, welcoming a greater number of NAS requires more of the school providers and the individual schools' organization and ways of working. It also requires a lot from the teachers, both in terms of knowledge and experience. In some parts of the country, it was already challenging to find teachers (and certified ones even more so) due to an already existing shortage in qualified teachers.

The OECD notes in their education policy outlook for Sweden that key policy issues include measures to convince teachers to stay in the profession, improve equity and comparability, and raise integration and learning outcomes for immigrants:

One of Sweden's main challenges is to continue improving the learning environment and teaching conditions, so that teachers feel more satisfied with their jobs and the most-qualified stay in the profession. At the same time, it is crucial to raise outcomes for immigrant students and continue to work toward quick integration. There is a need for a national framework to evaluate schools and teachers and ensure comparability of assessment results across schools. (2017, p. 4)

The increase of NAS after 2013 soon generated an entire department at the Swedish National Agency for Education created with the intention of helping equip schools and teachers to welcome and integrate migrant students into the Swedish school system. As a result, policy documents were launched as well as professional educational programs for teachers, together with teaching and diagnostic assessment material. Interest in researching the field also increased, but the Swedish Research Council's report from 2010 still notes a lack of research in the field, particularly regarding policy, organization and collaboration in terms of NAS. When it comes to equitable summative assessment among second-language learners, however, more needs to be done. Since assessment and grading are part of teachers' professional gatekeeping role and a passport to students' future opportunities, it is somewhat surprising that more research has not been presented in this area. How do subject matter teachers deal with novice language learners when their mission is to help all students reach predetermined goals, and where standards and objectives are based on a notion of all students sharing the same educational trajectory, when they clearly do not? Assessment is not an isolated part of education, but rather integrated into the educational practice as a record of what "goes in and comes out" of the education system. In the following, the aims and research questions of the present study are presented.

Aims and research questions

As noted above, big differences have been identified between schools in Sweden as regards how education and assessment are conducted with NAS. It was further stated that the big surge in the NAS population has brought challenges to a very decentralized education system. NAS in Grades 7 and up are at a disadvantage due to their late arrival. They are constantly behind due to fewer years of schooling compared to their peers, and have to learn a new language at the same time as they are supposed to cover the same course content in the subject matter courses and be measured on the same scale. This has appeared to be particularly problematic not just in history and social studies, but in all social and natural sciences in general, which are subjects with a big reading load (SOU 2017:54, p. 122). According to the OECD, the performance gap between immigrant and non-immigrant students in science is larger than the average across the OECD countries (OECD, 2015c). For NAS who arrive at the age of 16 or 17, it can be extremely difficult to reach the requirements to enter upper secondary school before turning 20. Referring to the NAE, Fratzke (2017, p. 20) mentions that 20% of the late-arriving students were able to enter an upper secondary program in 2013, a share that has dropped in succeeding years. Fratzke comments that this not only closes the door to upper secondary education, but also means a significant disadvantage when looking for employment, since a secondary school diploma is critical to labor-market success.

The aim of the present study is to contribute and gain some insights into the educational dimension of subject matter assessment in the named disciplines among language learners (i.e., NAS in Grades 7 and up), especially with regard to the following issues expressed by the OECD:

Entry to upper secondary and tertiary education depend on grades, but grading practices differ between schools. The grading- and upper secondary entry systems put heavy weight on fail grades [...] Teacher quality affects pupil outcomes, but teacher education is fragmented and of relatively low status in Sweden. (OECD, 2015b, p. 2)

The intent is to describe and examine the policies and practices of language use in relation to the *what* and *how* of assessment and hence its possible consequences for fairness in access opportunities and assessment outcomes. The varying practices and conditions across schools have been expressed in other contexts as well and are mentioned above. In 2009, the same concern was mentioned in a report from an investigation conducted by the NAE:

There is considerable variation in how schools conduct their work for the students with another mother tongue. The result of the study indicates a difference between whether the school has a higher or a lower proportion of students with another mother tongue. [...] From the equivalence angle, it cannot be fair that there is such variation between schools in the way teaching is organized. (2009, p. 25)

Teacher education is not targeted in this study, but teacher experiences, practices and beliefs can reflect both what works and what is needed in terms of teachers' professional development.

Research questions Study II:

- What are the subject content teachers' self-reported assessment beliefs and assessment practices regarding newly-arrived students?
- How are the teachers' self-reported assessment practices associated with their beliefs about the students and language policy in education (i.e., curriculum, syllabi, translanguaging practices and local organization)?
- Do teachers' assessment beliefs and assessment practices vary depending on school level and subject discipline? If so, *how* do they differ and *on what grounds*?

Conceptions of assessment and actual assessment practices have been found to be significantly associated (Brown, 2008). Therefore, teachers' self-reported assessment beliefs are believed to reveal something of the teachers' actual practices. In relation to this question, it is important to understand whether accommodations are used in assessment in order to reduce the impact of limited language. If so, what types of accommodation are used? Teachers' tacit knowledge is like a black box but can be made visible through self-reflection and self-developed assessment tasks or tests. Moreover, the data collection methods imply a mix of teachers' self-reported beliefs and practices as professed in questionnaires, semi-structured interviews and observations of teacher-developed assessment samples.

The second research question leans on claims made by Biggs and Tang (1999, p. 27) that desirable student learning depends both on student-based factors—ability, appropriate prior knowledge, clearly accessible new knowledge—and on the teaching context, which includes teacher responsibility, informed decision making and good management. As presented in Part 1 (p. 32), it is assumed, or strongly suggested, that teachers take the student into consideration when planning assessment and instruction, and therefore it is of interest to find out how teachers in the current context perceive their NAS. In relation to constructive alignment, it is also important to consider curricular goals and intended learning outcomes as expressed in syllabi. Given what was outlined in the introduction, the impact of language policy as found in translanguaging practices and local organization is also of relevance. Local organization also incorporates collegial collaboration and responsibility, in this case specifically for the NAS's development of the language of schooling to access and be able to perform in assessment situations. Is language part of instruction? If so, what type of language and how is it dealt with?

In the third research question, the comparative perspective, present in both Studies I and II, is in focus, including differences as well as similarities depending on teachers' context and disciplinary perspectives. Findings across disciplines and school levels can contribute to identify subject specific as well as generic issues connected with assessment regarding NAS and reveal if and in that case how school level and teachers' perceptions of the students' proficiency matter in assessment.

The next section provides a short review of recent publications and literature in relation to NAS, including Swedish research on education and migration.

Literature review and previous research

Newly arrived students' background and identity in research

According to a research review from the Swedish Research Council (Bunar, 2010), academic research as regards NAS is scarce in relation to many important issues such as policy analysis at a national level,

collaboration between different actors, instructional content and organizational models, to mention just a few. An exception is research on migrant children's language acquisition. Recent Swedish research in the field has especially addressed the newcomers' sense of identity, experiences of learning and inclusion (e.g., Nilsson Folke, 2017; Wernesjö, 2014), as well as the difficulties in integrating immigrant students into the Swedish education system due to a lack of cultural, social, educational and linguistic capital together with socio-economic limitations like class, gender and ethnicity (Sharif, 2017). In applied linguistics, a perspective often used in relation to second language learning, identity is perceived as central to understanding patterns in language learning and linguistic behaviors (Cummins, 2017, p. 160). Cummins (2017) argues that societal power relations affect the success of minority students and that there is a gap between educational policy and practice. The use of universal standards, curricula and national tests are considered to represent educational structures charged with values and ideologies that serve as tools for assimilation toward cultural and linguistic homogeneity. Identity negotiation is dichotomized with pedagogic efficiency. According to Cummins (2017, p. 169), these power relations need to be challenged and can be if schools communicate bilingualism as an advantage.

In the research review (2010), Bunar notes that a problem or deficiency perspective has been dominant in Swedish research. In descriptions of the migrant students' backgrounds in Sweden it is noted that there is a bigger proportion of NAS with mothers with a low educational level compared to migrants in other OECD countries, for instance (Cummins, 2017; OECD, 2014, 2015b), thus implying the possible negative effects a poor socioeconomic background may have on expected outcomes and academic progress.

Svensson Källberg (2018) sets out to critique the deficiency discourse while examining 15–16-year-old immigrant students' perceptions of their opportunities to learn mathematics. She notes how the students' perceptions of themselves are influenced by public discourse on segregation and migrants' supposedly deficient parents in comparison to Swedish students. However, policy documents intended to generate inclusion that speak of motivated migrant students who are using their strongest language to make fast progress also seem to contribute to exclusion if the students are unable to identify with the documents' descriptions, according to Svensson Källberg.

Bunar and Nilsson (2016, p. 399) maintain that "the pedagogical and social needs of newly-arrived students have enlivened politically and academically neglected questions of what constitutes appropriate organizational structures [and] pedagogical practices." If acted upon, the special needs of NAS consequently may contribute to the development of new pedagogy and have an effect on the whole education system. Below, challenges and recommendations in relation to teaching are presented via what has been found both in research and in the literature.

Teaching approaches and language across subjects in research

Whereas the aforementioned studies are based on a sociological framework, others take a more didactic perspective concerning classroom practices and migrant students' ability to take part in the meaning-making process in different subject content courses, depending on linguistic scaffolding or accommodations. Studies on language pedagogy, acquisition of Swedish as a second language and mother tongue instruction have been common research foci since the 1960s. Several studies examine teachers' perceptions of the role of language in their subject content courses not only in relation to NAS, but to migrant students in general, as outlined below.

Through classroom observations, Kouns (2010) investigates what opportunities the teaching strategies of a Grade 10 chemistry teacher may create for second-language learners to develop their Swedish language skills in the subject. While instruction appears to be very structured and expectations for the students are high, the language teaching strategies are unstructured, implicit and incidental. After looking at two Grade 10 physics teachers' language activities in class, Kouns (2014) concludes in a later study that the way scaffolding is developed in interaction with the students' oral and written performance is crucial for their ability to reach the knowledge requirements in physics. Using the oral and written production as a point of departure, teachers can assess the students in relation to the learning goals. The challenge for the teachers is to define which language elements to teach and to find time and space for scaffolding in class. Looking at teachers' language-oriented content teaching in a class on medicine in a vocational program at upper secondary school, Rubin (2019) notes that teachers need an opportunity to critically examine and reflect on their practice in order to integrate language-oriented teaching, and concludes that it is very demanding for individual teachers to change their teaching practices.

Ünsal (2017) addresses how language, gestures and physical artefacts are used in science class with emergent bilingual students who do not share the same minority language with their teacher or peers. She investigates how the use of various resources, such as gestures, and the use of the students' first language affects students' learning in science and concludes that valuing students' own ways of expressing scientific knowledge can help them learn the material. However, Karlsson (2019) notes that while translanguaging science classrooms give multilingual students greater opportunities to express themselves and participate in activities, there are areas in this practice that need development in order to create continuity in learning. All students need access to conceptual and linguistic tools, according to Karlsson, and if concepts and words are missing, the students need multimodal resources to clarify subject matter content.

In her study, Uddling (2019) focused on how language and content were processed in interaction with the students in so called "text talks", read alouds designed to construct meaning from decontextualized language. She examined the opportunities and limitations this offered the students. It was clear that the science textbooks impose great linguistic demands on the students. Whereas the text talks were found to offer opportunities for the students to develop academic registers, the teachers were not always clear about their purpose or implementation. Nygård Larsson (2011) looked at textbooks in biology and posits that the variety in academic discourse functions found in describing, explanatory and argumentative activities is demanding for students and teachers. She also notes that more challenging and advanced writing tasks are appreciated by the students and provide potential for learning. In her master's thesis, Nilsson (2015) finds that through an action research process the teachers become more aware of language as a resource for learning and for choosing activities in class. As a result, the teachers start to explicitly teach reading and writing strategies in biology.

Cummins (2017, p. 84) argues that while there is increasing language diversity in Swedish schools, teachers are not equipped to tackle it, which is especially problematic in the higher grades. The NAE (2015, p. 15) makes a list of significant needs in order to strengthen education among NAS, which includes more teachers with expertise in Swedish as a second language, more native language tutors and teachers, more adults in schools with native language expertise, skills development for all teachers in language development in all subjects, and digital remote solutions to ensure access to native language study guidance and teaching throughout Sweden. Below, language policy guidelines and suggested teaching models that are advocated in Sweden for the instruction of NAS are briefly presented.

Language policy guidelines and teaching models in literature and manuals

The OECD (2014) notes that language skills are an essential prerequisite for the ability of a foreign-born person to form networks with the native-born population and search for a job; knowledge of the host-country language is therefore a key factor in determining the speed and success of integration. The OECD further makes a point in relation to the difference between migrants coming to English-speaking countries and Sweden: whereas only 13% of migrants to Sweden in 2010 came from a country where Swedish is spoken, over 45% of migrants coming to Canada and 76% of migrants to New Zealand knew English. As regards language policy, Berglund (2017) comments in a report that unlike many other European countries, Sweden does not require proficiency in the national language. Whereas other countries take an assimilationist approach, Sweden has chosen *integration*, thereby advocating bilingualism and strengthening the migrants' mother tongues.

The "English only" immersion perspective which is found in many English-medium instruction contexts can be contrasted with a bilingual perspective, involving the systematic use of two languages, the target language and the first language. The two perspectives represent different language ideologies but also differing views on what is most effective for language learners. The first involves arguments that continuous exposure to target language instruction (in this case English) will help build academic skills quickly, so that students do not fall behind their peers. On the other hand, those in favor of two-language instructional programs posit that students will develop English more effectively if they can first develop literacy skills in their first language (Umansky et al., 2016). Cummins (2017, p. 150) sees Sweden as a pioneer in the latter, (the bilingual perspective), acknowledging the importance of students' L1 as a cognitive tool and for building students' identity, but he states that it has proved difficult in practice due to logistical and ideological issues. Regular instruction has not been reinvented, while at the same time the student body has become more multicultural and diverse.

Time and age are critical factors for the process of becoming functional in a language. Whereas everyday language is claimed to only take one to two years to develop, school language is said to take up to five years, but only if there is a lot of exposure, according to Cummins (1981, 2017). Age also has an effect: students who come during preschool need four years and students who come later need an additional two years (Cummins, 2017). In the meantime, use of the students' L1 is suggested. In manuals and documents from the NAE in support of teachers in general, but for the instruction of NAS in particular, the use of the students' strongest language is advocated (National Agency for Education, 2012b, 2016, 2018e, 2018g, 2012c, 2018i).

A common instructional model that is suggested to help scaffold language learners is called the teaching and learning cycle (TLC) (cf p.25) and was first developed by linguists and teacher educators in Australia (Callaghan et al., 1993; de Silva Joyce & Burns, 1999; MacArthur et al., 2008; Martin & Rothery, 1986). The model leans on Halliday's (1978) and Painter's (1999) research on children's acquisition of their L1 (Axelsson et al., 2006). Gibbons (2002) is often associated with the model, which contains four steps: modelling the text, guided practice, independent construction and building the context and field. The model implies starting with making the students' current and previous knowledge visible while using multiple activities to develop the language repertoire. The second step involves exposure to texts, examining the structure and identifying the linguistic characteristics of the genre in question. Thirdly, the students get to write their own texts in the same genre before discussing these together with the teacher, looking at structure, word choice and syntax, for instance. Lastly, the students are supposed to be in possession of enough knowledge and confidence to write their own texts (Axelsson et al., 2006; Gebhard, 2019; Gibbons, 2002, 2018). The TLC has roots in genre pedagogy and Vygotsky's thoughts on the need for dialogic teaching

with a more experienced adult, according to Axelsson et al. (2006). If applied in subject matter classrooms, this means that the content teacher should engage the students in explicit language instruction, which aligns with guidelines on how to work in linguistically-diverse classrooms. A governmental ordinance in 2016 required that government grants be given for the professional development of teachers in SSL methodology (SOU 2017:54). Some of the course material provided by the National Agency for Education is referred to below.

In summary, guidelines for instruction of NAS focus on two major features: letting students use their strongest languages, at least during the initial phase, and integrating explicit teaching of disciplinary language in the classroom. However, the questions that need to be resolved include when to switch from translanguaging to the use of the target language, how to align the students' language development with the language requirements of assessment, and what the content teacher's responsibility is in relation to the learning of different language registers. The next section provides a brief summary of what the guidelines say in a Swedish context (if indeed there are guidelines).

Subject literacy, language development and assessment

More research is needed in relation to assessment, but teaching manuals recommend the initial use of NAS's strongest language; not just in the classroom but for assessment as well (National Agency for Education, 2018i). In the governmental investigation from 2017, which was intended to improve equity in education for NAS (SOU 2017:54), key areas of development were identified, several of which refer to grading and assessment. It was stated that the mandatory diagnostic mapping of NAS is not used to a sufficient degree in order to be of help in the planning of instruction. In relation to the students, the individual knowledge progression of NAS needs to be made visible; they need to be presented with clear goals and students should be met with high expectations. In relation to learning and assessment, there must be continuous follow-up, a focus on oral discussions in class and oral presentations of knowledge. Moreover, there is a need for assessment guidelines below the E grading level in order to make learning visible until a passing grade has been reached. Teachers need professional development regarding how to work with NAS. It can be noted that inconsistent use of diagnostic mapping has been noted elsewhere as well (Cummins, 2017).

In course material for teachers provided by the NAE (2018i), the following summarizes the recommended approach to assessment of NAS:

- Use the student's collective linguistic resources
- Use varied forms of assessment, whatever works
- Use accommodations
- Formative feedback is important
- Peer feedback is advantageous and offers an opportunity for interaction
- A digital portfolio can be useful

As regards the students' language, it is argued that it would help if teachers were familiar with the characteristics of the different developmental stages in language acquisition, where faulty syntax, for instance, can be a good sign of "language in progress." It is noted that students' lack of subject concepts does not necessarily equal lack of content knowledge and vice versa; it was further suggested that

feedback must be selective in order to not overwhelm the students. Feedback should be based on the type of scaffolding the students have received and the language they need in order to perform the assignments (National Agency for Education, 2018i). Moreover, research shows that teachers often overestimate migrant students' language skills, focusing on decoding rather than understanding (Cummins, 2017, p. 73), which impacts assessment.

The OECD also seems to favor approaches involving the use of the students' strongest language. With reference to PISA and the reading test, the OECD (2019b) states that the use of the students' first language shrinks the performance gap between other students taking the test in their L1. As regards the national tests, it is argued that sources of bias in test grading need to be removed in order to create an objective benchmark (OECD, 2019a). The national tests were also pinpointed in the governmental investigation (SOU 2017:54), suggesting that students' use of their first language should be an option and thereby implying the availability of national tests in more languages, which would be a costly and difficult endeavor (today there are national tests in Swedish and English).

In summary, research dealing with newly-arrived students is multidisciplinary, emanating from linguistics, sociology, ethnology, economics, anthropology, pedagogy and political science. However, Bunar (2010) notes that cross-disciplinary research is still lacking and posits that the field would benefit from disciplines combining pedagogical and organizational perspectives. The latter is identified as an area in need of development so as to offer concrete advice to policy makers, according to Bunar. As already noted, little to no research has been conducted as regards teachers' assessment practices with NAS. In a Swedish context with a relatively long experience of second language learners, where syllabi in many subjects are language dense and the higher-order thinking skills in the subject matter learning objectives include language and discourse functions, it may seem there are broad cross-disciplinary experiences to draw from in research; this is what the present study intends to do.

Methods and material

This chapter presents the methods and material used in Study II. In order to answer the research questions, the aim was to provide as richly descriptive and explanatory results as possible. Different data collection procedures were employed to give a broad view of the field of study and provide complementary data, enabling a triangulation of the results. The theoretical lens and point of integration in the mixed methods' approach employed in Study II will be explained below. For Study I, see Part 2.

After a brief declaration of the theoretical perspectives of the study and description of the interactive mixed methods approach, the sampling process and the participants are presented, followed by a description of the instruments and the data collection procedures. Thereafter, the methods of analysis are outlined. The chapter concludes with a discussion of issues of validity, reliability and trustworthiness in relation to the study as well as ethical considerations.

The ontological and epistemological approach of the study

The present study is situated in the field of educational research, a field that is characterized by efforts to capture the complexity of the educational phenomenon. Ponce and Pagán-Maldonado (2014) note that research in the field has given rise to certain controversies, one of which refers to the dichotomization in defining the theoretical foundation of education, which is seen as either a cultural phenomenon or emerging from a natural phenomenon. According to the first perspective, educational research deals with abstract social phenomena and interpretations, whereas in the second, teaching and learning processes are the same regardless of context, and teachers' behavior can be regarded as causes and learning as effects (Ponce & Pagán-Maldonado, 2014). These perspectives can be identified as qualitative and quantitative research paradigms, respectively, thereby polarizing what is recognized as a hermeneutic versus a positivist stance; as argued below, however, educational research does not necessarily conform to either category. Alternatively, the researcher can take a pragmatic approach, focusing on how best to answer the research questions.

According to Smeyers (2006, p. 106), educational research should be characterized in terms of pluralism, since it must take into account societal developments as well as various principles, implying that the issue under investigation should be approached from a variety of theoretical stances, as is the case with the present study. Consequently, the research approach includes both inductive and deductive reasoning. In the inductive process, inferences are made from observations, providing a likely theoretical conclusion based on interpretation. The deductive approach goes from theory and rules to hypotheses, with the results expected to either confirm or nullify the theory (Bryman, 2016). As noted by Bryman (2016, p. 49), just as deduction contains a certain amount of induction, the inductive process includes features of deduction, since the researcher often has to move back and forth between theory and data in an iterative process; this is true of the present study (see examples below for an outline of the analytical process).

Smeyers (2006, p. 95) further argues that the educational field of study needs a different kind of research if it is to attempt to be true to the nature of what is being studied. In educational research, as well as other applied research (i.e., medicine, business and management research) aimed at solving practical issues affecting individuals or groups, voices have been raised for a need to make academic research more relevant to practice (Panda & Gupta, 2014; Pring, 2000a). In a discussion on the relevance of educational research, Bulterman-Bos (2008) argues that for the sake of validity, educational research needs to be grounded in teaching practice, with researchers who themselves are educators in order for

the research to be relevant. Labaree (2008, p. 422), in response to the same statement, infers that relevance is a tricky quality to define, since it is easier to recognize in retrospect. Instead, he advocates a need for barriers to be kept low to enable the exchange of perspectives and good conversation between practitioners in two areas (i.e., academia and schools). In the present study, the hope is that the researcher's background as an educator has had a positive impact on discerning the questions needed for research and thus will be able to contribute results that are of relevance to the educational context.

The stated aim of applied research is often to solve a specific practical issue that is affecting a group or an individual. The aim of the present study, to gain insights into assessment practices in multilingual schools, represents a practical issue that involves teachers, policymakers and curricula. As such, it affects both individuals and ultimately society, and therefore valid and equitable assessment practices are desirable at many levels. Investigating teachers' beliefs and practices can hopefully both reveal challenges that need to be addressed and present solutions that could be further explored and possibly exemplified in policy.

The theoretical underpinnings to a multilayered study such as the present are cross-disciplinary and complex. Whereas teacher beliefs and students' language development lean on socio-constructivist epistemology and a dialectic ontology, building on interaction with the surrounding society, the normative foundation of target language as expressed in curricula builds on an essentialist ontology with a structuralist epistemology, characterized by the identification of named existing languages. The foundation of construct validity theory, as found in assessment, is also at an intersection of positivist and realist thinking, depending on the intended construct of assessment (Slaney, 2012). From a positivist and operationalist stance, constructs are viewed simply as descriptive summaries of particular classes of behavior implied by the content, according to Slaney, whereas a realist perspective implies that test performance is seen as reflecting the effects of objectively existing entities. On the one hand, "constructs" can be presumed to be observable behaviors (i.e., a realist stance), while on the other hand, constructs are perceived as objectively existing but unobservable entities (i.e., a positivist stance) (2012, p. 3). Slaney notes that different theorists either take a stand with one of the two or a combination of both. As Fulcher and Davidson (2007, p. 11) infer, "validity theory occupies an uncomfortable philosophical space in which the relationship between theory and evidence is sometimes unclear and messy." Messick (1989, p. 23) gives voice to a shift in epistemology, stating that previous validity categorizations offer a useful guide for disciplined thinking about the process of validation but cannot serve as the prescriptive validation model to the exclusion of other approaches. It was from this realization (i.e., that we see from our beliefs, which are not watertight categories since they change because of observation) that post-positivistic validity theory evolved, together with Messick's unified validity framework, advocating the integrated validity category, namely, construct-related evidence (Fulcher & Davidson, 2007). It is on this notion of the integrated and consequential basis of validity that the present study rests.

The perspective of selecting methods that best address the research questions stems from pragmatic foundations and "what works" (Cohen et al., 2011; Johnson & Onwuegbuzie, 2004), as noted above. Mixed methods have been considered particularly useful in research taking practice as their point of departure. From a methodological perspective, a pragmatic approach, including the use of several complementary data collection instruments, as found in mixed methods, was perceived to be useful for the purposes and questions of the present study, with the aim of attaining some degree of breadth as well as depth in data and findings. A discussion of the strengths and weakness of mixed methods methodology follows below.

Mixed methods design

Mixed methods research has been defined as one of three methodological or research paradigms (Johnson et al., 2007, p. 113). It is sometimes referred to as the “third research paradigm” and portrayed at the center of a continuum, where “pure qualitative” is found at one end and “pure quantitative” at the other (cf. Johnson et al., 2007). Johnson and Onwuegbuzie (2004, p. 15) note that the goal with mixed methods research is “not to replace either of these approaches [qualitative or quantitative] but rather draw from the strengths and weaknesses of both.” Reinhardt and Cook (1979) argue that there is no reason for researchers to be constrained by either one of the traditional methodological paradigms, since these are largely arbitrary. Smeyers (2006), suggesting a complementary perspective, posits that quantitative research alone fails to offer reasons for human actions, in the same way that qualitative research alone fails to problematize and offer alternative conclusions. Greene (2007) and Greene et al. (1989) identify five reasons for mixing: *triangulation*, *complementarity*, *development*, *initiation* and *expansion*. Since then, more purposes of mixing have been distinguished; Bryman (2016), for instance, posits that combining the two approaches makes the results more useful to practitioners.

Criticisms of mixed methods have been raised, stemming from a confusion as to what is being mixed and whether mixing is thus possible or fruitful. Two different kinds of criticism can be identified. One comes from a perceived *incommensurability perspective*, stating that paradigms cannot be united in the same investigation since positivist and interpretivist theories cannot be used to study the same phenomenon. Quantitative and qualitative paradigms represent different, conflicting worldviews that refer to particular ontological, epistemological and axiological positions (Sandelowski, 2000; D. Scott, 2007; Varaki et al., 2015). The conflicting ontological theories either believe the world can be recognized and described as it is (i.e., a quantitative paradigm), or the world is constructed by individuals or groups of individuals (i.e., a qualitative paradigm) (Pring, 2000a, 2000b). The challenges are thus rooted in theoretical and philosophical assumptions, which Pring (2000b) refers to as a false dualism.

A second perspective of criticism echoes the argument of the paradigm fallacy, stating that mixed methods’ *raison d’être* as a third paradigm leans on an enforced separation between methodologies and their practitioners that has grown out of two stereotypes. Gorard (2007) argues that methods should not have been divided at the outset. It is further maintained that methods cannot be qualitative or quantitative, since questionnaires and interviews alike can be used both in a quantitative and a qualitative paradigm. Data from open-ended methods, which would traditionally be categorized as qualitative, are indeed quantifiable, and conversely, data from close-ended methods, which would traditionally be labelled as quantitative, do not need to be quantified (Allwood, 2007; Åsberg et al., 2011; Symonds & Gorard, 2008). The latter perspective can be perceived as a more post-modern stance, which implies taking an anti-epistemological position, rejecting fixed “modes of thinking” that are associated with modernism (Potter, 2017, p. 144). Allwood (2007) refers to an old categorization between the natural and the social sciences from the 19th century that is outdated, rooted in old beliefs where the qualitative research approach was interested in the “nature” of the object under study, whereas the quantitative approach only asked “how many” and “how much.” Allwood notes that today both approaches are interested in explaining and describing phenomena.

Regardless of stance, whether indeed a separation must be made to stay within paradigms, or to remove paradigms altogether, a need has been articulated to clarify what is meant by method in relation to the label mixed methods. Cohen et al. (2011) state that distinguishing methods from methodology implies making a distinction between *ways of collecting data* versus *style of research*. The decision as to which instrument (i.e., method) to use often follows from a decision as to which kind of research to conduct.

Onwuegbuzie and Leech (2005, p. 382) suggest that the terms quantitative and qualitative should be replaced by confirmatory and exploratory research, arguing that not all quantitative approaches are positivist or all qualitative, hermeneutic. The data for this study were collected using both quantitative and qualitative methods (i.e., ways of collecting data) for reasons of complementarity and to gain a broad as well as in-depth understanding of the overarching research question: *What are teachers' assessment beliefs and practices among newly-arrived students?* The aim was to provide both descriptive and explanatory results. One possible method for investigating teachers' assessment practices is observing their actual behavior in practice, looking at assessment procedures and interaction in the classroom or assessment samples used in the courses. Another possible approach is to address and ask teachers, either in an interview or a questionnaire, bearing in mind that teachers may state what they believe to be appropriate practice rather than presenting what they actually do (Hofer, 2002). For the purposes of this study, a survey, interviews and assessment samples were used. The survey was used to gather quantitative data about teachers' beliefs in multiple contexts, and the subsequent qualitative interviews aimed at providing thicker descriptions from a sample of teachers together with their assessment samples. To quote Guest and Fleming (2015, p. 582), “for the purposes of this [study], qualitative data refers to nonnumerical data [...]. Conversely, quantitative data are numerical in nature and can be in various forms—dichotomous, ordinal, interval”; in the present study, the quantitative data are mainly ordinal (see section below on the instruments). The non-numerical data in this case are interview data transcribed into written text, written assessment samples and open-ended questions in the survey.

Out of concerns for academic rigor, validity and trustworthiness in mixed methods research, a need for defensible and transparent research designs has been articulated (Johnson et al., 2007; Schoonenboom & Johnson, 2017); as a result, certain more or less fixed designs have been proposed. The present study is framed by an *interactive mixed methods design* (see Figure 3.1 below), which can be described as both sequential and concurrent. The interactive approach has been interpreted as follows for the current study: Phase 1, containing the data collection procedures, is sequential, since the quantitative instrument preceded the qualitative (the participants in the survey were asked if they would consider sharing more of their experiences in a follow up interview, to which they were asked to bring assessment samples). The first overview of the survey data informed the consecutive design of the interview guide. Phase 2 consisted of data analyses and was more integrated in character, thus creating an interactive process with concurrent analyses of interview data and survey data. In the first step, the two data sets were organized, transcribed, thematized and analyzed individually, followed by a comparison of results and emerging themes, and then a return to separate analyses. In Phase 3, findings were triangulated and interpreted in relation to the research questions; see figure below:

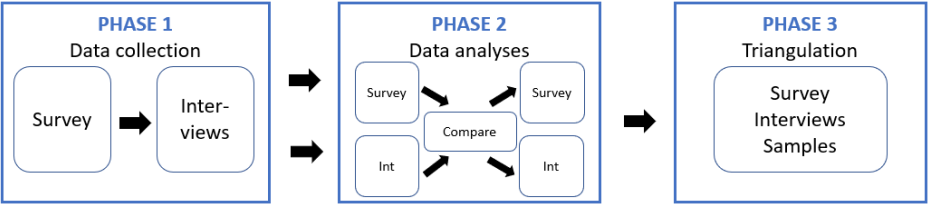


Figure 3.1 Interactive mixed methods' design

Interactive mixed methods approaches are sometimes used in contrast with typological approaches. Whereas the latter have a more fixed design, the interactive approach is described as a dynamic process. Based on Maxwell and Loomis (2003), Schoonenboom and Johnson note the following:

In an interactive approach, unlike the typological approach, design is viewed as an interactive process in which the components are continually compared during the research study to each other and adapted to each other. (2017, p. 121)

As noted in the quote above, the interactive design implies a continuous comparison and analyses throughout the process. Hence, the data collection, which was preceded by literature reviews and development of the questionnaire, leading up to the conceptual framework, also consists of analyses. The timing and point of integration is not indicated in Maxwell's (2013) description; however, Maxwell and Loomis (2003) distinguish five key components that the researcher should check regularly to guard the fit between them and deliver a proper fit in the end product. The five components of the interactive design are: *goals*, *conceptual framework*, *research questions*, *methods* and *validity*. In the present study, the five components have been compared and adjusted regularly, ending with a triangulation and description of the final conclusions, as outlined in the discussion that follows the findings.

Below, the sampling process and participant characteristics in Study II are presented, starting with a description of how individuals were identified for participation, including sampling approach and sampling constraints. An account of response rates and estimates of the likelihood of achieving population representation (i.e., the extent to which the participants can be presumed to reflect a defined population) follow, succeeded by a reporting of descriptive background data.

Sampling process and participants

When deciding sampling size, "factors such as expense, time and accessibility frequently prevent researchers from gaining information from the whole population" (Cohen et al., 2011, p. 143). In the present study, which aimed at collecting both quantitative and qualitative data, a tentative minimum sample was set at around 200 for the initial survey, although a bigger sample would have been welcomed. According to recommendations, a sample size of 30 is recommended to conduct some form of statistical analysis (Cohen et al., 2011), and the purpose was to create room for several subsamples ($n = 30$) within the total for comparison across groups. Time was an important factor when making a final decision as to the amount of the total sample ($N = 196$) for the investigation in progress. The target population was teachers in the named disciplines in schools with an increase in number of NAS after 2015. For the semi-structured interviews, a tentative sample size was set at around 12, equal to the final sample size for Study I. However, a suitable sample size for the interviews was determined based on what could be accessed and the teachers' subject identity ($N = 13$). Below, the sampling process and the participants in Study II are outlined.

Purposive sampling

Purposive sampling is used to focus on specific unique issues or cases in order to access those who have in-depth knowledge about particular issues due to expertise or experience (Cohen et al., 2011, p. 157). The target group for this study was teachers in the natural and the social sciences with experience of NAS and second-language learners at the critical ages of 13 to 19, thus representing students from Grades 7 to 12. The overarching reason for this was that both the natural and social sciences have been

identified as demanding subject areas for language learners due to the complexity and language density of the texts (SOU 2017:54). The chosen age group was targeted since NAS at this level receive grades, while at the same time they have little time to learn both the language and the subject content. Hence, teachers in subjects within the social and the natural sciences were invited to participate, and not just biology and history as in Study I; the reason for this was both practical and logical due to the structure of education and similarities between subjects. Moreover, the aim was to find a larger sample for the survey. In Sweden, instruction in the social sciences (i.e., geography, history, religion and social studies/civics) can be integrated up to Grade 9 under the umbrella of the social sciences. In the natural sciences, teachers often teach several of the subjects in the natural science domain (i.e., biology, chemistry, physics and technology). When it comes to grading, however, individual subject grades must be assigned (National Agency for Education, 2020b). Moreover, in research as well as in the educational literature, the subjects are often grouped together in discussions on subject literacy. Consequently, distinguishing between disciplinary differences within the natural or social sciences was not of interest for the purposes of this study; instead, the focus was on shared interdisciplinary features in related sciences.

As regards experience with NAS, the planned target groups were: (a) teachers with more extensive but recent experience with NAS in municipalities that had welcomed the largest proportions of migrant students after 2015; as well as (b) teachers with potentially longer experience in suburban areas. The purpose was to facilitate comparison between teachers with shorter versus longer experiences of language learners to find out if (and in that case how) exposure and experience would impact teachers' assessment beliefs and practices. Initially, Grades 7 to 9 were chosen, since NAS can be found in any regular class at this level, as opposed to upper secondary, where the NAS are dispersed into different programs, making it more difficult to track teachers with experience with NAS. After consulting statistical data from the Swedish Migration Agency and the NAE, a sample was determined using lists of municipalities receiving the largest numbers of migrants after 2015; the belief was that contacting schools with Grades 7 to 9 would generate enough teachers to produce a sample of at least 200 teachers, including teachers from one or two major cities or suburban areas. At first, 10 municipalities were selected from among those with the highest proportion of NAS in the relevant age group. A list was made of all schools in the selected municipalities offering education for the relevant grade levels. Contact information was gathered for principals and in some cases for administrators at the municipal level or at the schools, when contact information for principals was not available. In the first round, an email was sent to principals; this, however, only generated three responses where personal references to a person at the NAE could be used.

At this point, the introductory program was added to increase the sample, as well as Grades 10 to 12 in upper secondary. Since the target group was contexts with late newcomers, it was already clear that not many students would be found at national programs, since entry is only granted after receiving enough passing grades after Grade 9 (see Part 1 on education in Sweden). Students who are not yet eligible for national programs are found in introductory programs. For the final sample, a total of 108 municipalities were contacted, representing all the initial small towns with large recent influxes of NAS as well as mid-sized towns with both recent and longer experiences of receiving migrants, plus a sample of schools in larger cities.

The Swedish Association of Local Authorities and Regions (SALAR) provides a classification of all municipalities in Sweden that has been used for this study. The classification contains three main groups—A, B and C—and nine subgroups. In Table 3.1 below, the structural parameters for the municipal categories are summarized.

Table 3.1 *Classification of Swedish municipalities (according to SALAR, 2017)*

	Definition of the municipalities used for this study/First part of SALAR's definition	The rest of SALAR's definition and structural parameters Population and commuting patterns
A	Large cities	... and municipalities near large cities Population of at least 200 000 More than 40% commute to work in a larger city
B	Medium sized towns	... and municipalities near medium-sized towns Population of at least 50 000, and 40 000 in the urban area Commuting municipalities near medium-sized towns, more than 40 % of the working population commute to work Commuting municipalities with a low commuting rate near medium-sized towns – municipalities, less than 40 % of the working population commute to work
C	Smaller towns	...urban areas and rural municipalities Commuting municipalities near small towns, more than 30 % commute to work Rural municipalities with a population of less than 15 000 inhabitants in the largest urban area, very low commuting rate (less than 30 %)

The first column represents the first part of SALAR's definition, which has been used for this study's sample. The second column presents the rest of SALAR's definition, including structural parameters that determine the categorization

After using statistics to identify municipalities and creating the chart referred to in this study (cf. Figure 1.14, p. 49), it came to my knowledge that a different classification had been used: SALAR changed the categories in 2017. However, when comparing old and new categories, the differences were small. Moreover, creating a categorical variable for municipalities in the analysis did not generate any actual differences between groups. As will be seen below when the analytical process is described, comparison of teacher beliefs and practices based on municipality was dropped.

Convenience sampling

Although a purposive sampling was used initially, accessibility to the target groups was not granted due to the use of principals as gatekeepers. This resulted in a convenience sampling, or as it is sometimes called, opportunity or accidental sampling (Cohen et al., 2011, p. 155).

At the schools where principals or administrators had responded positively and agreed to ask teachers to participate, a total number of 635–700 teachers got the link, according to reports from principals, and thus an opportunity to respond. The span depends on varying reports from the principals, as explained below. As seen in Table 3.2, 196 teachers submitted their answers to the web questionnaire: 7 in June and 189 in August to December.

Table 3.2 *Overview of response rate in the web-survey*

Number of contacts (i.e. principals/ administrators)	Dissemination of link	Number of dropouts	Number of respondents
536	635–700 teachers	70	196

According to the response rate figures, 70 teachers did not complete the questionnaire and did not submit their answers (i.e., 26% of the 266 teachers who registered to answer the initial questions), leading to the conclusion that the questionnaire was too extensive. The reports from the pilot had indicated the same, and therefore the questionnaire was shortened and reassessed. The new questionnaire was estimated to take 20 minutes. However, some questions may have required too much time from ambitious respondents, especially the optional open-ended comment section at the end of several of the items. The aim was a minimum of 200 respondents, and the scope of the sample was continuously widened during the data collection process to ensure a bigger sample (see Table 3.6). Contact with teachers was made via principals or school administrators, who acted as gatekeepers. The principals were given two options: (a) to ask teachers in the target group for volunteers to participate and then have me send them the link; or (b) come back with a list of email addresses. Since this seemed to lessen the odds of participation, another option was offered: they could have the link to the questionnaire immediately after agreeing to send me a response mail as regards the spread of the link. Several months and a couple of email reminders later, some of them still had not responded or could not remember the number of teachers. Others claimed to have shared the link with all of their staff—up to 30 teachers—and not just the targeted teacher groups. All of these factors represent a constraint concerning the sample.

The table below gives an overview of the residence of the participants in the survey. Among 196 respondents, 187 stated their place of residency, leaving 9 incomplete/missing responses that were included in the analyses.

Table 3.3 *Stated cases of municipalities and number of respondents in the survey (exact numbers)*

	A	B	C	Total
Municipalities	5	19	38	62
Respondents	20	78	89	187

Out of the 108 contacted municipalities (cf. Table 3.6), representing 37% of Sweden's 290 municipalities, at least 62 municipalities participated in the survey, as shown by Table 3.3 above. This means that 20% of Sweden's municipalities were involved (albeit on a small scale, in some cases with a single representative). The largest proportion is found in category C, which coincides with the main target group. Category B includes some smaller municipalities with large proportions of NAS between 2015 and 2017, but according to the classification, they are categorized as B due to commuting patterns. Category B thus includes both medium-sized towns and smaller municipalities near medium-sized towns, as seen in Table 3.1.

Participants in the survey and population representativeness

Below, descriptive background data are provided as regards the participants in the survey. Comparisons are made with statistical data from the NAE from the same school years (2018f, 2019d) and Statistics Sweden (SCB in Swedish) (2017) to show population representativeness (i.e., to what extent the study participants can be presumed to reflect the defined population).

The number of participants in the survey was 196. Participants included 110 women and 82 men (4 other or missing), meaning that 56% of the respondents were women and 44% men aged between 26 and 67 (men: $M = 44.24$, $SD = 10.69$; women: $M = 45.48$, $SD = 9.09$). The average age was 44.97 years ($SD = 9.8$). Teachers' mean age depends on the school level, but the national statistics (Statistics Sweden, 2017) mention 46 years, which is close to the mean of this study. The national gender distribution at upper secondary was rather similar—52% women and 48% men—but this depends on the subject. In this study, the proportion of the social sciences and languages exceeds that of the natural sciences (see table below). In compulsory school, Grades 1–9, the gender distribution is skewed, with 75% women, which probably is more evenly distributed in Grades 7–9. The average teaching experience was 14.71 years ($SD = 8.5$), and the range was 1 to 38 years. The national average was 12 to 13 years, depending on school level, and was slightly higher at the upper secondary level, indicating that the participants in the survey had slightly longer teaching experience than the national mean.

Swedish teacher certification ($n = 181$) was held by 93% of the participants, which is slightly higher than the national average. According to reports from the NAE, fewer and fewer teachers are certified: 81.4% are certified in upper secondary, but in subjects such as history, chemistry, biology, social studies and religion, 85%–90% are certified, which may explain the frequencies in this study. On the other hand, the participants in this study were selected by their principals, and therefore uncertified teachers may have been excluded. At the introductory program level, fewer teachers are certified; if they are included in the count for the upper secondary level, the proportion decreases from 81.4% to 78.5%, which, however, is higher than the year before, when it was 76.8%. In compulsory public schools, the number of certified teachers in Grades 1–9 is 71.5%. The frequency varies depending on subject, level and provider and is higher in public schools and in Grades 7–9.

The collapsed frequency of teachers at the three school levels in the study was $n = 114$ for Grades 7–9, $n = 55$ for the introductory program and $n = 47$ for Grades 10–12. A few ($n = 5$) mentioned *other*, mostly represented by preparatory language classes and Swedish for immigrants (SFI). However, teachers often teach at more than one level, as illustrated in Table 3.4 below. A majority of the teachers in this study teach Grades 7–9; this proportion is representative of the expected distribution, since the target group was teachers in the natural and social sciences with experience with NAS. Teachers in upper secondary (Grades 10–12) have fewer newcomers in their classes, since NAS tend to spend more time in the introductory program or are often found in different vocational programs or education for adults. If teachers go between levels, it is more common for teachers to combine the introductory program and Grades 10–12 than the introductory program and Grades 7–9 in this sample. This is logical for organizational reasons, since the introductory program usually is located in upper secondary schools. For pedagogical reasons, it might have been reasonable to believe that the frequency of teachers from Grades 7–9 would be higher, since the introductory program shares the same syllabus and both measure attainment in relation to Grade 9 objectives. (For the distribution of teachers across school levels and subjects, see Appendix H.)

Reports concerning teacher education indicated $n = 115$ for Grades 7–9, $n = 99$ for Grades 10–12 and $n = 21$ for other. Education level frequency and teachers working in Grades 7–9 seem to correlate, whereas 99 report being educated to teach upper secondary (i.e., Grades 10–12) but only 47 teachers work at that school level. No teacher education is geared toward the introductory program, and when looking at teacher educational background in this group, 35% ($n = 19$) are educated to teach Grades 7–9, 44% ($n = 24$) to teach Grades 10–12, 20% ($n = 11$) to teach a combination of both, and one participant chose the category *other*. When asked about work experience, 74% ($n = 144$) reported having other professional experience.

The disciplinary distribution of the teachers shows a slightly higher frequency of teachers in the social sciences. Teachers could report disciplinary combinations, including language subjects and subjects outside the natural and social sciences, and a higher number of the social science teachers reportedly teach and are certified to teach both a content subject and a language. Language in this case could mean English, Swedish or Swedish as a second language (SSL). Language was included in the questionnaire, since experience of second language learning was perceived as having a potential effect on teaching/assessment approaches in the discipline, which was the topic of investigation.

Table 3.4 below shows the frequencies of subjects taught and certification in the subjects. The frequencies seen in the table reveal slightly higher numbers for teachers who teach the subjects than those certified, especially in the natural sciences (NS), where 83 were teachers but only 66 were certified. The non-response rate ($n = 28$) is higher in relation to certification, which might suggest a lower inclination to respond if non-certified. The frequency number of language teachers ($n = 10$) includes teachers who are certified to teach in the NS and SS disciplines but currently only teach languages, 1 teacher who teaches language and math and consequently was not included in the NS category, 1 teacher who reported language and “other subject” (i.e., psychology), and 2 teachers who reported currently teaching Swedish for immigrants. However, all of these indicated having experience in the topic of the study and were thus kept in the sample.

The teachers were asked about their language competence and experiences of other languages. Ninety-eight percent claimed to be proficient, native or near-native speakers of Swedish ($n = 188$) and 93%, proficient English speakers ($n = 171$). A total of 6 teachers reported having knowledge of Arabic, ranging from basic to advanced skills. Knowledge in other languages comprises 19 different languages, mostly represented by basic skills. The reported languages in order of frequency were: German ($n = 32$); Spanish ($n = 13$); French ($n = 10$); Finnish ($n = 6$); Serbo-Croatian ($n = 5$); Greek ($n = 3$); Bosnian ($n = 2$); Danish ($n = 2$); Syrian ($n = 2$); Russian ($n = 2$); and a number of languages with only one representative (Albanian, Armenian, Czech, Hungarian, Norwegian, Persian, Polish, Turkish and sign language). The purpose of the question was to find out about their general language learning experiences, a factor that could be an area of further research in a future investigation.

Table 3.4 below shows the frequency distribution (exact numbers and valid percent) of the teachers’ background variables expressed as school level, subjects taught, certification, age (mean), teaching experience (mean), gender and the most frequent languages. (For the list of less frequent languages, see above.) For the languages, the total percent is reported with the valid percent in parentheses (see Table 3.4 continued), since the valid frequency based on the total number who responded gives a slightly skewed picture (e.g., 3 out of a total of 8 who responded that they know Arabic). Valid percent only relates to the respondents who chose to name the language in question instead of the total population. Stated language competence seemed more relevant in relation to the total population in the survey, and not only in relation to the internal distribution for every specific language.

Table 3.4 Teacher background variables

Variable	Frequency	Valid Percent
School level		
<i>Grades 7–9</i>	108	55
<i>Introductory program</i>	35	18
<i>Grades 7–9 & Intro</i>	3	2
<i>Grades 10–12</i>	28	14
<i>Grades 7–9 & 10–12</i>	2	1
<i>Intro & 10–12</i>	16	8
<i>Grades 7–9 & Intro & 10–12</i>	1	0.5
<i>Other</i>	2	1
<i>Total</i>	195	100
<i>Missing</i>	1	
Subjects taught		
<i>Natural sciences (NS)</i>	83	44
<i>Social sciences (SS)</i>	56	29
<i>Language</i>	10	5
<i>NS & language</i>	2	1
<i>SS & language</i>	38	20
<i>NS & SS & language</i>	2	1
<i>Total</i>	191	100
<i>Missing</i>	5	
Certified to teach per subject/subject group		
<i>Natural sciences (NS)</i>	66	39
<i>Social sciences (SS)</i>	52	31
<i>Language</i>	11	6.5
<i>NS & language</i>	2	1.2
<i>SS & language</i>	34	2.2
<i>NS & SS & language</i>	3	1.8
<i>Total</i>	168	100
<i>Missing</i>	28	
Age	44.97 (Mean)	
Teaching experience	14.71 (Mean)	
Gender		
<i>Female</i>	110	56.4
<i>Male</i>	82	42.1
<i>Other/decline to report</i>	3	1.5
<i>Total</i>	195	
<i>Missing</i>	1	

Table 3.4 Teacher background variables (continued)

Variable	Frequency	Total percent*
Languages		
<i>Swedish advanced</i>	183	93 (97)
<i>Swedish intermediate</i>	4	2 (2)
<i>Swedish basic</i>	1	0.5 (0.5)
<i>English advanced</i>	70	36 (41)
<i>English intermediate</i>	89	45 (52)
<i>English basic</i>	12	6 (7)
<i>Arabic advanced</i>	3	1.5 (37.5)
<i>Arabic intermediate</i>	1	0.5 (12.5)
<i>Arabic basic</i>	4	2.0 (50)

*Valid percent in parenthesis, see explanation above.

In the section below, descriptive data as regards the interviewed teachers are provided. As already noted, the teachers who volunteered for the interview all took part in the initial survey.

Participants in the semi-structured interviews

One of the questions in the questionnaire asked the respondents to state if they would be interested in a follow-up interview. Twenty-one of the 196 respondents agreed to be interviewed and left their contact information, which was either an email address or, in some cases, a phone number. A first selection was made to contact teachers representing an even spread across subjects and school levels, which can be said to represent a purposive sampling. A couple of teachers had only marked SSL as their subject taught and were not considered to be in the target group. A certain number of teachers did not respond, and one dropped out after the first initial contact. The teacher in question wanted to be interviewed by phone or Skype and withdrew when asked to meet for the interview. In the next step, more teachers were added, representing a convenience sampling. The first 11 interviews were conducted face-to-face at the schools. The last two were conducted over Skype and phone for practical and technical reasons, the last on speakerphone since Skype failed. The table below provides an overview of the participants and background data. The selection of background data for the table represents key features for comparison. The term *school level* that was used previously is here defined as *grades taught*.

Table 3.5 below presents data concerning age, gender, years of teaching experience, years of teaching NAS, school level/grades taught, subjects taught and certification in the subjects, which is marked with an asterisk. Finally, the type of town is described (cf. Table 3.1 with municipality categorization).

Table 3.5 Main background features of the interviewed teachers

	Age	Gender	Teaching experience (years)	Experience NAS (years)	Grades taught	Subjects taught * =certified	City/town
T1	67	Male	18	4–5	Intro, 10–12	Social Studies,* (History*), (Religion)	Medium-sized town
T2	52	Male	6	6+	10–12	Physics,* (Chemistry*), (Math*)	Medium-sized town
T3	36	Female	10	6+	7–9	Geography, History, Religion, Social Studies, Swedish 2,* (Spanish*)	Medium-sized town
T4	36	Male	8	2	Intro, (10–12)	Geography,* (History*)	Medium-sized town
T5	49	Female	15	4–5	Intro, (7–9)	Geography,* History,* Religion,* Social Studies*	Small town
T6	30	Male	3	2–3	7–9	Biology,* Chemistry,* (Physics, Math, Natural science*)	Medium-sized town
T7	51	Male	21	2–3	10–12	Physics,* (Math*)	Medium-sized town
T8	46	Male	10	2–3	Intro, 10–12	Religion,* Social Studies,* (Swedish 2, Geography, History, Math)	Large city
T9	32	Female	4	4–5	7–9	Geography, History,* Religion, Social Studies, (Swedish 2*)	Large city
T10	52	Female	11	4–5	Intro, 10–12	Religion,* Swedish 2*	Medium-sized town
T11	38	Female	18	2–3	7–9	(Geography,* History, Religion, Social Studies*), Swedish 2	Small town
T12	36	Female	10	6+	7–9	(Geography, History,* Religion, Social Studies, Swedish*) Swedish 2,* English*	Medium-sized town
T13	28	Male	1	1	7–9	Social Studies,* Geography, English	Medium-sized town

Parentheses around subjects indicate that these are subjects that are not taught this school year or at the time of the interview but have been taught previously.

Whereas some of the teachers teach NAS or language learners in the introductory program (T1, T4, T5) or in a special language class (T3), others teach regular classes with some integration of NAS (T2, T6, T7, T9, T10, T13). Three of the interviewed teachers have a different situation. One (T8) used to teach in the language introductory program from 2015 to 2018, but now teaches social studies and religion at national programs with no NAS, and therefore makes retrospective reflections about differences and experiences in relation to both. Another teacher (T11) taught social studies in regular classes and in a language class with NAS during the peak, 2016–2018, but now teaches in a small group with special needs students. T11 is now a mentor to a non-certified colleague who teaches the social sciences in Grades 7–9, and therefore relates conversations they have had about assessment in the colleague's classes. Another teacher (T12) has a similar background and refers to her experiences from last year, when she had NAS in her social studies classes; this year, she only teaches language courses.

Among the teachers participating in the interviews, 2 had a doctoral degree in one of their subject disciplines, 1 had written a master's thesis on assessment, and 3 had been appointed at their schools to be involved in special arrangements for NAS or students with disabilities. The same teachers plus 1 had either been elected "lead teachers," serving as a mentor to colleagues, or did part time projects at the municipal level (e.g., working toward digitalization or computer assisted learning); another had worked as a teacher trainer, supervising student teachers. One was engaged in issues of integration as a politician, and 1 had personal experience of being a non-native language learner of Swedish. The group was self-selected or assigned by a school leader to answer the questionnaire, and therefore the sample cannot be claimed to be representative of an average group.

The instruments

The questionnaire

In the following, the design of the data collection instruments is described before looking at the data collection procedure and the types of data. The first draft of the web-based questionnaire was developed during the late fall of 2017 using the web-based software Survey&Report. Several drafts later, after consultations with experts in the field of survey development and language development among second language learners, the survey was piloted on a group of teachers in the target group during the spring of 2018. After their comments, the questionnaire was revised before the final version was distributed for the first time in June 2018 (see below for more about the collection procedure).

The questionnaire includes a broad selection of independent background variables to provide answers to an initially broader intended research question: If, and if so how, teacher background and context has an effect on teachers' assessment beliefs and practices among newly-arrived students. The question was later modified to only cover subject background and school level for several reasons: extent and scope of data for the purposes of the present thesis and time at hand. Several background variables needed recoding and collapsing cases and the first two background variables had already generated a lot of interesting results. However, several background factors were run through SPSS for the initial analyses and will be used for another study.

The background variables provide descriptive data concerning subjects taught, school type, length of teaching experience (both overall and specifically in relation to second language learners), number of NAS, and languages represented in class; in addition, data were gathered on teachers' socio-demographic characteristics, such as age and gender. Other background variables include teacher education and professional development and background languages of the teachers. As already noted,

the main background features relating to subject discipline and school level, as expressed in the third research question, were used for the analyses. Other contextual factors covered include policy as regards collegial collaboration, language use, mapping of the NAS’s background knowledge and the use of national tests. However, these questions represent dependent variables and are used in comparative analyses between teachers representing different subjects and school types.

The dependent variables focus on how teachers perceive their students’ ability in relation to oral and written language skills, the verbal activities in the subjects’ knowledge requirements and teachers’ self-reported practices concerning instruction and assessment. Responsibility refers to perceived policy and organization as regards language instruction and accommodations for the students in relation to assessment accessibility. See Figure 3.2:

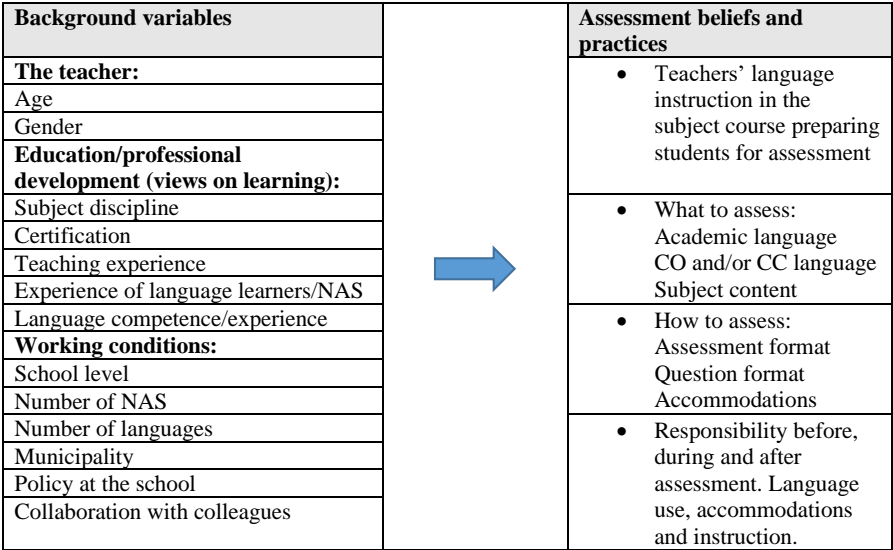


Figure 3.2 Factors influencing teachers’ assessment beliefs and practices

Questions were thus developed to generate background data about the teachers, their professional profile, and experience both of teaching in general and of language learners and NAS in particular. The purpose of some other questions was to generate information about the organization of the schools in order to gain an understanding of the teachers’ working conditions. The questionnaire is composed of 43 questions, plus 3 additional questions that the respondents could fill out in order to: share some important knowledge on how to work with NAS; leave their contact information, if they were interested in an interview; and comment on the questionnaire. Out of the total of 46 items, 18 used a nominal scale, 15 an ordinal scale, 2 a ratio scale, and 11 were open question format. For the items on the ordinal scale, measures of frequency were used (e.g., *always, often, sometimes, rarely, never*) or scales measuring perceived importance (*very important, fairly important, not important*) or difficulty (*difficult, fairly difficult, neither easy nor difficult, fairly easy, easy*). Thirty of the nominal and ordinal items had an open comment box in case the respondents wanted to leave an additional comment. See Appendix C for the full questionnaire.

The interview guide

A semi-structured interview is usually characterized by a situation where the interviewer has a number of general questions organized in a question scheme and the order of the questions can alternate (Bryman, 2016, p. 260). A certain structure is needed for comparability across interviews. Hence, an interview guide was developed based on the research questions and the questionnaire for the semi-structured interviews in the shape of a checklist. The themes were as follows:

- Teacher background
- Teaching context/organization
 - Number of NAS, number of languages
 - Collaboration with colleagues
 - Local policy
- Assessment/samples
 - What:
Construct, content/language
 - How:
Test administration
Scoring/accommodations
 - And:
Student performances
 - Experiences: what works and what does not
 - Preparation and follow-up in class
- Views on language
 - Language approach in teaching
 - Language approach in assessment

Below is a description of how the instruments were used, providing a presentation of the data collection procedure and the collected data types were generated by the study.

Data collection procedure and data types

The design and preparation of Study II started in the fall of 2017 with a given time frame of two years; this was due to the fact that a licentiate thesis had already been completed and defended in 2015, which counted as half a doctoral dissertation (i.e., Study I, which was summarized in Part 2 of this thesis). However, additional time was granted due to parallel work at the NAE and teaching in the university department.

The results of the present study build on teachers' self-reported perceptions and practices, as in Study I. As described previously, the first step in the data collection was a large-scale survey and the second step

consisted of the semi-structured interviews and collection of assessment samples. Below, the data collection process and data types are presented.

The survey

The initial plan was to collect data before summer 2018, with teachers’ experiences from the current school year fresh in mind. Email requests were sent to some principals, asking for their consent to contact teachers, but very few (four municipalities with whom I had a personal connection via a contact at the NAE) responded. Since the final revision of the questionnaire took place in mid-May, the initial goal was abandoned, and the new plan was to disseminate the questionnaire in August. When I spoke to principals on the phone, they believed the survey probably would generate better response rates in August, since teachers are tired in June. However, a couple of schools/districts had already notified teachers and wanted to let them participate before summer, and therefore a decision was made to do both.

The first round in June did not generate many answers, but they were added to the total after the second round during the fall of 2018. The survey closed in December 2018, after the response time had been extended three times. Since some of the items asked for teachers’ experiences from the previous semester/school year, I concluded that the questionnaire data needed to be gathered by no later than the fall of 2018. Due to poor response rates, the time frame was extended as much as possible. Table 3.6 below provides an overview of the sample invited to participate, time of contact, municipal category, and number of schools. In one case, a contact was made on the municipal level and the person in question did not come back with information on the number of schools that had been invited in that municipality. A total of 536 principals or administrators were contacted (cf. Table 3.2), and some had to be contacted repeatedly; this is not reflected by the frequencies in the table. The purpose of the table below is to show the distribution across municipalities during the sampling process.

***Table 3.6** Number of schools and municipalities contacted for the survey by month*

	A Large city	B Medium-sized town	C Smaller town	Number of schools
June	1	2	8	11
August	-	6	14	34
September	4	11	20	152
October	3	13	20	174
November	-	2	4	8
Total:	8	34	66	379

The semi-structured interviews

The semi-structured interviews were conducted between April and June 2019. The teachers who had shared their contact information in the survey, stating that they would be interested in participating in an interview, were contacted via email starting in February 2019. A time was decided for a phone call when the letter of consent was presented, and details were discussed concerning the time of interview. The interviews were pushed toward the end of the semester, since several teachers had national tests and other projects that needed to be finished. It also took some time to coordinate the trips across the country. On the phone call and when meeting for the interview, the purpose of the study was once again outlined;

most of the participants were not interested in a lengthy presentation, however, stating that they remembered the information from the survey. I introduced myself briefly and related the previous research study and work experience, which was also done in the letter sent before the survey. Via an email before the meeting, the teachers were asked to bring some assessment samples for the interviews (see Appendix F-G for the email request).

The interviews lasted from 39 minutes (T8) to 1 hour and 18 minutes (T5), generating a total of 756 minutes or 12.5 hours of interview data (see Table 3.7). Teachers are labelled T1 to T13; T stands for teacher, and the number was assigned according to the order of the interviews.

Table 3.7 *Duration of interviews with teachers in minutes*

T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13
50	69	46	68	78	65	60	39	48	55	61	70	47

The assessment samples

To complement their self-reported assessment beliefs, the teachers were asked if they would share some written assessment samples from their subject disciplines to be discussed during the interview (e.g., tests, instructions for individual written/oral projects, lab reports or group assignments). The teachers were informed that the assessment samples would be used at the interview to discuss, for example, the following:

- What was assessed in the assignment/exam?
- How did the assessment work for the newly-arrived students?
- Were accommodations used during the time of the assessment? If so, which ones?
- How did it work for you as a teacher to assess the assignment? Experiences/insights?

As in Study I, the amount and character of the assessment samples shared by the teachers varied (see Table 3.8 below).

Table 3.8 *Teachers' assessment samples*

Teacher	Assessment samples
T1	Coursebook with sample questions used for assessment
T2	2 written exams
T3	1 written exam, 1 written assignment, 1 group assignment
T4	3 written exams
T5	4 written exams, 1 written/oral assignment
T6	2 written exams
T7	2 written exams
T8	2 written exams, 1 written assignment, 1 oral group assignment
T9	4 written exams
T10	3 written exams, 1 written/oral group assignment
T11	1 written quiz/home assignment
T12	-
T13	3 written exams, 1 national diagnosis, 1 written assignment

Most teachers brought assessment samples for the interview. In a couple of cases, test samples were attached to emails after the interview. In one case, questions and instructions were pasted into an email, thereby presenting the questions but not the actual test. Some teachers also provided examples of student performances on the actual assessments, which were discussed briefly at the interview. When asked how a substitute teacher, for instance, would know how to use and grade the tests, a couple of teachers stated that they have assessment manuals, but most teachers referred to the knowledge requirements in the syllabi and tacit knowledge as to how scoring and grading is performed. However, several acknowledged that it would not be easy for an outsider to know how to grade the exams. Hence, as outlined below, the assignments/exams were used together with teacher reports from the interviews in the analyses.

Analyses of data

In this section, a brief account will be given of the methods used for the analyses of data (i.e., the survey, interview and assessment samples). The analytical treatment of the statistical data is described first, followed by a presentation of the approaches used for the qualitative data.

Descriptive statistics and quantitative analyses

For the analyses of the survey data, the statistical software SPSS (Statistical Package for Social Studies) was used. The decision as to which statistical test to use is dependent on the kinds of data. Since most of the items on the questionnaire were nominal or ordinal (see section above on the questionnaire), the data is considered non-parametric (Cohen et al., 2011). In the quantitative analyses, both descriptive and

inferential statistics were employed to describe and compare teachers' beliefs with a check for effect sizes between different combinations of variables in non-parametric analyses of variance.

Descriptive statistics summarize data and are presented in text, charts and tables, including frequencies and, for continuous variables such as age or years of teaching, mean and range. Different latent independent variables were tried in bivariate and multivariate analyses to explore the explanatory value they might possess on possible differences in the dependent variables on teachers' assessment beliefs and practices, language policy, subject matter language instruction and beliefs about the students. To describe the interrelationships between two categorical variables, cross-tabulations were used.

Non-parametric tests were run in SPSS using independent samples to find differences between sub-groups on different dependent variables. The Mann-Whitney test of significance was used for two independent groups and the Kruskal-Wallis test of significance for three or more independent groups. The Mann-Whitney test is based on ranks, comparing the number of times a score from one of the samples is ranked higher than a score from the other sample (Bryman & Cramer, 1990), thereby overcoming the problem with chi-square in relation to low cell frequencies. According to Henriksson (2008), the Mann-Whitney provides one of the strongest measures, including for small samples. The Kruskal-Wallis operates in a similar way to the Mann-Whitney, but enables us to see if there are differences between three or more groups (Cohen et al., 2011, p. 659). In order to identify where the significant differences were to be found, the Bonferroni post hoc test was used. The level of significance in all tests was set at $\alpha = 0.05$.

The results are reported in tables, and include mean rank, Mann-Whitney U, and asymptotic significance (2-sided) for the bivariate analyses between subjects. For the multivariate analyses between school levels, the Kruskal-Wallis only states whether there is a significant difference between school forms, not where it can be found. Here, the cross-tabulations provide information, which is presented in the text. All results from the non-parametric tests, both significant and non-significant, are presented in Appendices I–K.

When national statistics could be found, comparisons were made between the survey sample and the national teacher population for the year the survey was conducted (2018) in order to see how the participants in the investigation correlate with the general teacher population at the time.

Qualitative content analyses of interviews, surveys and assessment samples

For the purpose of exploring the participants' beliefs regarding assessment among NAS, thematic content analysis was used for the interview data and the open comment sections in the survey. Content analysis defines the process of summarizing and reporting written data, the main contents and their messages (Cohen et al., 2011, p. 563). Thematic analysis (TA) shares history with content analysis and they have often been used interchangeably (Braun et al., 2019). However, Braun et al. note that TA has increasingly been recognized as an "approach in its own right," or rather as an umbrella term for several approaches. The term "theme" can be conceptualized as a pattern of shared meaning organized around a core concept or idea, according to Braun et al. (2019). Using their typology for TA, the procedure used for this study could be defined as codebook TA, where some themes are determined in advance of the analysis and themes can be conceptualized as domain summaries. Other themes are often based on data collection tools and responses to certain questions, summarizing what participants said in relation to a topic or issue. As such, they can imply multiple or even contradictory meaning and content, according to Braun and Clark (2019).

The preparation of the qualitative analyses started with transcribing all the interview material. Since the focus of the analyses was on content and not language, very basic transcription conventions were employed, (descriptions are found on p. ix), and only literal statements were transcribed. The interview data were then read several times, themes were highlighted, useful segments and examples were translated into English, and they were collected in separate documents and coded and organized thematically for use in the analyses. As expressed by Nowell et al. (2017, p. 6), during coding, researchers identify important sections of text and attach labels to index them as they relate to a theme or issue in the data; the purpose of this is to identify both emerging peripheral themes and themes evolving from the use of the interview guide. Consequently, both deductive and inductive processes were employed: deductive, since known categories from the survey and interview guide were used to create a provisional template (King, 2004), including teachers' perceptions of the students, perceptions of language and perceptions of assessment, and inductive, as two new categories were identified—*Comparison NAS/Swedish native students*, and *Instruction, other (approach, working methods)*—directed by the content of the data.

The open questions in the survey were examined and compared with the interview data to find similar confirmatory or deviating contradictory beliefs. Complementary data was selected and organized in tables to be used for the presentation of the findings.

The assessment samples were anonymized and coded according to the label of the contributing teacher. Overall descriptions were made as regards subject discipline, thematic content and number of questions. In the next step, more detailed descriptions were made (similar to test specifications), including question formats, modalities (inclusion of pictures, diagrams), scaffolding (inclusion of wordlists, startup sentences), and scoring. In the third step, the assessment samples were analyzed in relation to syllabi, academic function words/cognitive discourse functions, higher- versus lower-order thinking skills, and Cummins' matrix. The next step of the analyses implied mixing, as outlined below.

Mixed analysis and triangulation

After the independent parallel analyses of the quantitative and qualitative data, the material was compared and triangulated in relation to the research questions and the identified themes. A template was made with three parallel columns representing and organizing the teachers' reports from the interviews together with descriptions of the tests and the test items. The first column includes statements of what is being assessed and the second reports how assessment is performed. The third column contains summaries and examples of the teachers' assessment samples as an illustration to enable comparison of teacher beliefs and teacher practices. Findings from the survey were presented in tables and figures, and data were interpreted and compared across datasets. Validation through triangulation was made and results were presented in the text.

Summary

To sum up, Table 3.9 below presents an overview of the analytical process; table inspired by Käck (2019).

Table 3.9 Overview of handling of data and analytical procedures

<i>Data procedures</i>	Quantitative	Qualitative	Mixed analysis
1. <i>Prepared</i>	The data from Survey&Report were transferred into SPSS	All interviews were transcribed. Open-ended answers from the survey were read and some examples were transferred to Word. Assessment samples were coded.	Both sets were independently analyzed and compared.
2. <i>Explored</i>	The data were visually inspected and processed, and items were recoded and transformed when needed. Elaboration of alternative models of independent and dependent variables.	The data were read several times to get an overview. Thematic categories were identified and segments of categorized data were translated into English. Assessment samples were described.	Both sets of data were explored in parallel.
3. <i>Analyzed</i>	Descriptive data, frequencies (percentage), cross-tabulations, chi square to find significance between different combinations of independent and dependent variables. Non-parametric tests were conducted (Kruskal-Wallis, Mann-Whitney U) for significance.	The data were analyzed using the analytical tools: the language matrix and the assessment continuum. Similarities and differences were identified as well as teacher profiles.	Both sets of data were compared in relation to the research questions and teacher categories/independent variables.
4. <i>Represented</i>	The results were presented in tables and figures connected to descriptive analyses.	Thematic results and teacher profiles were presented using interview data and assessment samples.	Quantified qualitative data in tables or text, and qualitative quotations connected to quantitative results.
5. <i>Interpreted</i>	The data were interpreted and explained with respect to the research questions and related to literature and previous research.	The results were compared to theories and literature and discussed in relation to the research questions.	Interpretation of the combined results in relation to the research questions was made.
6. <i>Validated and interpreted</i>	Tests of significance were conducted, and results were interpreted.	Validation through triangulation.	Validation through triangulation.

Validity, reliability and trustworthiness

As noted above concerning mixed methods, *validity* in research is a quality indicator that requires defensible and transparent designs aimed at explaining the research questions. Validity can take many forms and relates to design and methods as well as findings. Validity means that the results represent the phenomena I set out to investigate and can be sustained by data. However, there are specific validity requirements that differ depending on methodological stance. Validity in quantitative research can be improved through careful sampling and appropriate statistical treatments. In qualitative research, *understanding* or *accuracy* have been proposed as alternatives, since the researcher is responsible for the meaning given to the data and the inferences that are made (Cohen et al., 2011, p. 180f). In mixed methods research, it has been argued that the term validity should be replaced by *legitimization* (Onwuegbuzie & Johnson, 2006) due to the integration of what is perceived of as sometimes antagonistic canons of validity stemming from different research paradigms.

Reliability relates to features of accuracy, consistency, replicability and precision. Cohen et al. (2011, p. 199) note that for research to be reliable, it must demonstrate that if it were to be carried out on a similar group of respondents in a similar context, similar results would be found. The most important factor in producing reliable results is to be clear, critical and honest in evaluating sources of error (Swedish Research Council, 2017). An evaluation of this study's potential sources of error is presented in the limitations section following the discussion of the findings.

In qualitative methods, the term *trustworthiness* is used to parallel the conventional quantitative quality assessment criteria of validity and reliability. Lincoln and Guba (1985) refined the concept of trustworthiness by defining four key criteria: *credibility*, *transferability*, *dependability* and *confirmability*. When conducting data analysis, the researcher becomes the instrument for analysis and is thus responsible for disclosing the methods of analysis with enough detail to enable the reader to determine whether the process is credible (Nowell et al., 2017). Lincoln and Guba (1985) maintained that data collection triangulation would enhance credibility. By applying consistent methods and providing detailed thick descriptions, the researcher can enable transferability, described above as replicability in relation to reliability (generalizability is a term that is used in a similar vein). In quantitative research, generalizability generally relates to how well the sampling mirrors a larger population. In qualitative research, generalizability can also refer to transferability of results and theory (i.e., not only methods), where readers can make use of the research findings in their own contexts (Akker, 1999; Denzin & Lincoln, 1994). In this study, the aim is to provide as detailed and transparent descriptions as possible to facilitate transferability at all levels.

The third criterium listed above—dependability—means that the research process is logical, traceable, and clearly documented (Tobin & Begley, 2004). This has been a goal throughout this study, with the collection, evaluation and improvement of different types of documentation during the analytical stages. Dependability equally includes an aspect of consistency and stability of findings over time, according to Korstjens and Moser (2018), referring to Lincoln and Guba (1985), which time will prove in a basic research study as the present. To conclude, I believe trustworthiness comes out of honesty, transparency, and a conscientious, humble approach, acknowledging that validity should be seen as a matter of degree, not an absolute state: “Hence at best we strive to minimize invalidity and maximize validity” (Cohen et al., 2011, p. 179).

Ethical considerations

The Swedish Research Council (2017) posits that “ethical considerations in research are largely a matter of finding a reasonable balance between various interests that are all legitimate.” Hence, various disciplines and methodological stances accentuate different concerns. Smeyers and Depaepe (2010), for instance, express skepticism as regards overdue emphasis on statistical data in educational research, stating that it can be useful for professional interests but devastating in its consequences for schools and society. In qualitative research, Bridges (2002, p. 72) points to an ethical dilemma when a teacher becomes a researcher investigating his or her own community, as was exemplified previously in relation to my role as a former teacher: “the insider researcher will always be something of an outsider in his or her own community by virtue of becoming a researcher.” In qualitative research, the researcher’s “effect” on both participants and as an interpreter implies a need for constant self-reflexivity (R. Berger, 2015) concerning the researcher’s intentions, stance and possible bias. Although I presented myself as a researcher, my background as a teacher and director of education at the National Agency for Education is stated on my University website. At the interviews I shared briefly about my language teacher background and work at the NAE. The teachers were invited to inquire more about my background before proceeding, but they were focusing on what they were there to share. One teacher seemed interested in my connection with the NAE and expressed a need for more guidelines and assessment material.

Since this study uses quantitative and qualitative data, I have considered all concerns expressed above as I have worked on how to handle data and present my findings (i.e., in relation to my background as a teacher and at the NAE, my personal understanding and interpretation of findings, including my translation process from Swedish to English which is not always straightforward, and the ability to generalize the results from my sample).

Balancing different interests sometimes implies conflicting demands from society, the research community and ethical boards. Ethics thus relate to my motives as a researcher and use of research, as much as the actual process and how participants and data are handled. For this study, the ethical guidelines of the Swedish Research Council (2017) were followed. In order to fulfil the information requirement, the participants in the survey, as well as the principals who were contacted, received information about the purpose of the study. The teachers who volunteered for the interview were informed that the data collection would be documented by audio recordings. All participants were asked to sign a consent form, which was collected before the interviews (see Appendices D-E). The requirement of consent stipulates that participants have the right to withdraw their consent at any time, which was stated in the consent form. Since participation was voluntary and all the participants are adults, no ethical vetting was considered necessary for this study. Moreover, all efforts have been made to conceal the identities of the individual participants in the study. The teachers were thus assigned a T for teacher, together with a number. Potential ethical issues were considered throughout the study. To summarize, I will quote myself from a previous text (Reierstam, 2018): “Theory gives ethics its initial meaning, but then ethics have to do with how my principles or standards are transformed into action in specific situations, ‘to walk the talk.’”

In the next section, the main findings of Study II are presented in the same order as the data collection and in relation to the themes of the research questions and the analytical tools. The main results from the survey are introduced, followed by a description of the results of the semi-structured interviews together with examples from the assessment samples.

Findings in Study II

The findings in Study II are based on aggregated results from quantitative and qualitative investigations, merging results from the survey, the semi-structured interviews and the assessment samples. In the following, the results are presented thematically aiming to provide explanatory and descriptive data in response to the first overarching research question:

- *What are the subject content teachers' self-reported assessment beliefs and assessment practices regarding newly-arrived students?*

The aim of the second research question is to find explanatory connections between teachers' reported assessment practices and surrounding factors:

- *How are the teachers' self-reported assessment practices associated with their beliefs about the students and language policy in education (i.e., curriculum, syllabi, translanguaging practices and local organization)?*

The results in relation to this question are presented in six thematic sections based on the structure in RQ 2 above and constructive alignment, starting with teachers' perceptions of the students, moving on to language in the intended curriculum (as expressed in the national curriculum and syllabi) and the implemented curriculum (in relation to possible translanguaging practices and local organization), before finally connecting all the components with assessment (cf. constructive alignment as illustrated in Figure 1.7, p. 32 and Table 1.8, p. 44).

Significant results in relation to the third research question, aiming at investigating the effect of independent background variables on teachers' assessment beliefs, are discussed more in depth in section five below comparing teacher groups:

- *Do teachers' assessment beliefs and assessment practices vary depending on school level and subject discipline. If so, how do they differ and on what grounds?*

However, comparisons across teacher groups are also integrated into the other thematic sections to describe, juxtapose and analyze the teachers' assessment beliefs.

As described above, the results are presented in six thematic sections, starting with a brief background as to the teachers' experiences with NAS and describing the language composition of the groups. The second section presents the teachers' perceptions of the students and perceived prerequisites to be assessed in the subject content courses. The aim of these first two sections is to provide an understanding of the context that is believed to contribute to shaping the participating teachers' beliefs about NAS.

How language is perceived and dealt with in the subject courses as well as local language policy for NAS at the schools is the topic of the third section, since beliefs about language and language use constitute an integral part of assessment, as expressed in the research question. Teacher beliefs about assessment are presented in the fourth section. In section five, results are presented showing significant differences found in the non-parametric tests between teachers in the natural and social sciences, between teachers in the social sciences with or without a language teaching combination, and between teachers at different school levels. All results from the non-parametric tests, significant as well as non-significant, are presented in Appendices I–L. The reporting of results concludes with a brief overview of teachers' expressed needs as regards professional development in the fields of assessment and students' language development.

Data from the survey and the semi-structured interviews are presented in synthesis, but survey data comes first before looking at individual teachers’ comments from the interviews, presented in excerpts where *T 1* (etc.) represents the teacher and *R* the researcher. (Transcription conventions can be found on p. ix). Descriptive statistics are presented in tables, in some instances in bar charts, and commented upon in the text. Significant results from the non-parametric tests (section five) are shaded with blue in the tables. Frequencies are in valid percent.

Sometimes, a survey question is quoted in connection to the results, while at other times it is only referred to as *Q* followed by the item number (e.g., *Q24*). The questionnaire is attached in Appendix C; note that it is written in Swedish, as the survey was conducted in Swedish. Findings from the assessment samples are presented in section three to provide illustrative examples of the interviewed teachers’ assessment practices.

The six sections are labelled as outlined to the left below, with subheadings as listed to the right:

Language composition of the groups and teachers’ experiences with NAS	<i>Teachers’ experiences with NAS</i> <i>Average number of L2 students and teachers’ previous experience with language learners</i>
Teachers’ perceptions of the students	<i>Students’ language proficiency</i> <i>Perceived difficulty in different language skills and effects on grades</i> <i>Age, varying schoolasic background, perception of knowledge and students’ progress</i> <i>Lack of cultural cues and everyday experiences</i>
Teachers’ experiences and beliefs about language in education	<i>Disciplinary language taught and used in class</i> <i>When language is taught</i> <i>Language use in class and translanguaging</i> <i>Organization and collaboration—who is responsible for language?</i>

Teachers’ assessment beliefs and practices across subjects and schools

- The “what” of assessment and the intended learning outcomes*
- Challenges in relation to language and higher-order skills*
- Challenges in relation to accountability and students’ memorization strategy*
- The “how” of assessment—the use and importance of different assessment types*
- The “how” of assessment—local language policy and accommodations*
- The “how” of assessment—teachers’ language policy and accommodations*
- The “how” of assessment—examples from the interviews and the assessment samples*

Comparison across teacher groups

- Comparing NS and SS teachers*
- Comparing SS and SS + language teachers*
- Comparing teachers across school levels*

Professional development in assessment and language teaching methods

(no subheadings)

After the six thematic units, a discussion with concluding remarks frames the investigation in relation to the first overarching research question on teachers’ assessment beliefs and practices regarding NAS. Before looking at the findings, the tables below present an overview of the interviewed teachers’ labels (T1, T2, etc.), and serve as a reminder of the teachers’ subject disciplines and school levels as outlined previously in Table 3.5. The distribution is not evenly spread across subjects, which was discussed in the section on the sampling process in the methods and material section.

Table 3.10 The subject disciplines of the interviewed teachers

Natural sciences	Social sciences	Language (SSL, EFL)
T2, T6, T7	T1, T3,* T4, T5, T8, T9, T10, T11, T12, T13	T3, T9, T10, T11,* T12, T13

*see explanation below

Table 3.11 *The school levels of the interviewed teachers*

Grades 7–9	Introductory program/Language class	Grades 10–12
T3, T6, T9, T11, T12, T13	T1, T4, T5, (T3), (T11), (T12)	T2, T7, T8, T10

Teachers T3, T9, T10, T11, T12 and T13 appear in both the social sciences and language columns in Table 3.10. The asterisks for T3 and T11 indicates that they are not certified in this discipline, while the others are certified to teach both social science subjects and a language (SSL and/or EFL). T3 is a certified language teacher, recruited to teach the social sciences in a language class. Teacher T3 teaches one class, Grade 8, that represents a specific kind of language class. It consists only of NAS and students with a different L1 than Swedish. Due to the teacher’s profile, experience and proficiency as an SSL teacher, she was assigned a class where she teaches languages and the social sciences. T11 is certified in the social sciences but has taught SSL in a language class for NAS and now assists a non-certified colleague in the social sciences. T12 is certified in the social sciences, SSL and EFL but teaches only in the language class this year; it is a preparatory class where the students gradually participate in mainstream education, with the goal of eventually being fully integrated into a regular class. The parentheses in Table 3.11 indicates that these teachers are considered to be teachers in Grades 7–9 due to the focus of the subject matter courses in this study. For an overview of the combined distribution of subjects and school levels of the participants in the survey, see the description of the participants in the methods and material section and Appendix H.

Language composition of the groups and teachers’ experiences with NAS

In order to understand the teachers’ conditions at the schools and the teaching context, one of the questions in the survey asked about the number of languages and second-language learners in the classes: *(Q13) Approximately how many languages in total were represented among the newly-arrived students in your classes last year?* The question had four alternatives: 1–2, 3–5, 6–9 and 10 or more. According to the frequency analysis, the most common number of languages was 3–5, covering more than half of the respondents and representing 58% of the total ($N = 194$); see Figure 3.3 below.

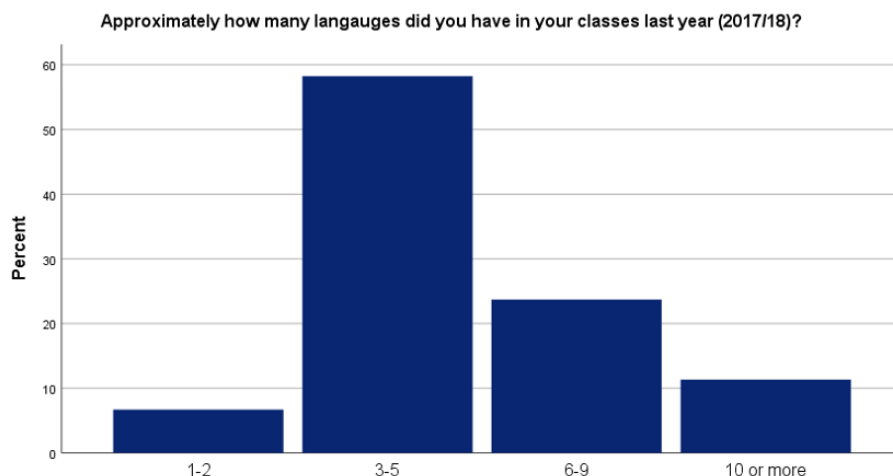


Figure 3.3 Teachers' appreciation of the average number of languages in their classes

The total number of languages differs depending on school level, but the second category (3–5 languages) was the most frequent across all levels. Comparing school types showed that for Grades 7–9 ($n = 80$), the distribution resembled the total frequency analysis, as shown in the bar chart above: 61% reported 3–5 languages and 18% reported 6–9. The distribution between the other groups was quite even, with approximately 10% in each. For the introductory program ($n = 55$), the category 3–5 languages was the most common as well. For regular programs in upper secondary school, Grades 10–12 ($n = 46$), 63% reported 3–5 languages and 28% reported 6–9 languages. The upper secondary teachers formed the only group reporting 0% in the category 1–2 languages, which means that the participants from that school level represent a sample with larger language diversity (i.e., 3 languages or more).

Teachers experiences with NAS

As regards the teachers' self-reported experiences with language learners and multilingualism, more than one-third of the teachers had fairly long experience. In answer to question *Q11—How long have you had newly-arrived students in your classes?*—35% ($n = 65$) responded 6 years or more, which dates back to before the peak in NAS in 2015; 29% ($n = 54$) responded 4–5 years of experience and collapsing the 1 year and 2–3 years categories shows that slightly more than one-third (36%, $n = 66$), had 1–3 years of experience. Here, a comparison between municipalities was made to find out if the teachers' reports as regards their length of experience of NAS are related to the recent increase in NAS in smaller towns and rural areas. See Table 3.12.

Table 3.12 Teachers' reported length of experience with NAS in their classes (%)

Municipality type:	1 year	2-3 years	4-5 years	6 years or more
A	5	35	30	30
B	9	33	26	32
C	10	18	32	39

As the frequencies show, a similar pattern can be found regardless of municipality type. Most teachers have at least 2 years' experience or more, although one possible explanation may be that teachers with less experience declined to participate or that only teachers with more experience were asked by their principals to participate. Category C, which represents smaller towns, including municipalities with a sudden increase in NAS after 2015, actually seems to have a fairly large group of teachers with more than 4 years' experience and the largest group with more than 6 years' experience, although teachers who were involved in the 2015/16 school year when the surge in NAS was peaking probably can be found in the category 4–5 years.

Average number of L2 students and teachers' previous experience with language learners

In Q14, the teachers were also asked to report the average number of Swedish L2 speakers who had a different mother tongue than Swedish in their classes: *State how many of your students were second-language learners in your groups last year (2017/18) and had a different mother tongue than Swedish.* Since the number varies across school levels and between classes at the same school, it is difficult to estimate an average. However, in order to get a picture of the proportion of second language learners, including NAS,³ the teachers were asked to estimate and report for each school level relevant to them. The teachers had the options *a few, one-fourth, half, three-fourths* or *most of the students*. Whereas most, if not all, of the students in the introductory program have a different mother tongue, most teachers reported that they had a few NAS/language learners in each class, up to a maximum of one-fourth of their students. In Grades 7–9, it is noteworthy that 12% ($n = 23$) answered *most of the students*, which may imply that these are so-called language preparatory classes, although this is not revealed by the data.

To the question regarding whether the teachers had any prior experience with language learners before receiving NAS (Q12), a majority responded yes, even if the occurrences of NAS may have been random. According to the frequency analysis, the number of teachers who had second-language learners in their classes *all the time* represented 30% of the total ($N = 189$). Additionally, 29 % responded *now and then*, implying that almost 60% have had more frequent exposure and experience with language learners, whereas only 14% responded *never* and 3%, *don't know* (see figure below).

³ Since teachers in the piloting process reported that it was difficult to single out NAS due to the distinction "4 years or less", the broader category of second language learners was used.

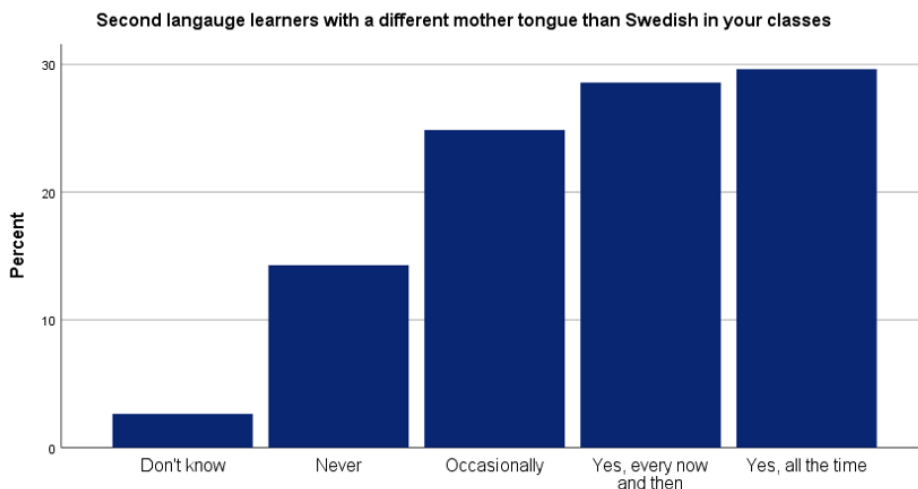


Figure 3.4 Teachers' experience of language learners prior to receiving NAS

Many smaller municipalities with a recent surge in NAS were included in the survey to find out how they deal with supposedly new exposure to second-language learners. However, as seen by the results, a majority of the participants in the survey did not seem unfamiliar with NAS or second-language learners in their classes. Ninety-five percent ($n = 186$) of the respondents stated the municipality in which they worked, with the distribution between the three categories (cf. Table 3.3) showing that 11% worked in a large city, 42% in a medium-sized town/commuting area and 48% in smaller towns (a slightly skewed distribution). The target group was teachers in smaller towns with a recent surge (post-2015) of newly-arrived migrant students. The challenge was to get teachers in large cities and suburban areas with supposedly longer experience with multilingualism and language learners to participate in order to enable comparison. (Several principals declined since they had already been invited to participate in other investigations.) The uneven distribution of NAS across smaller and sometimes remote towns since 2015, as mentioned in Part I, has implied a sudden and important change toward more multicultural and heterogeneous student bodies in many small municipalities. This uneven distribution did not have a significant effect on the results of this study.

In summary, the participants in the survey work in a context where they generally are faced with 3–5 different languages, represented by an average of a few students in each class, and they are fairly used to second-language learners. Given the participating teachers' frequent exposure to language learners and experience with NAS, questions arise as to how they perceive the students and their abilities and challenges using Swedish monolingual syllabi. Teacher reports from the survey and the interviews provide some insights in the following.

Teachers' perceptions of the students

In Swedish schools, so called "early skills mapping" (a diagnosis of the NAS's literacy skills and previous schooling) has been mandatory since 2016 in order to facilitate placement of students by

assessing their current academic level (see Part 1 on education in Sweden). The results from the diagnosis are supposed to serve as guidance when planning instruction for NAS and for the allocation of time and resources between subjects. In the survey, the teachers were asked if they had taken note of results from the general literacy assessment in steps 1 and 2 of the diagnosis, and whether they had experience of performing the subject content diagnosis in step 3, which can be administered by content teachers (Q23). A majority responded no to both questions, which is consistent with other findings in relation to the same topic (Mörtlund, 2020).

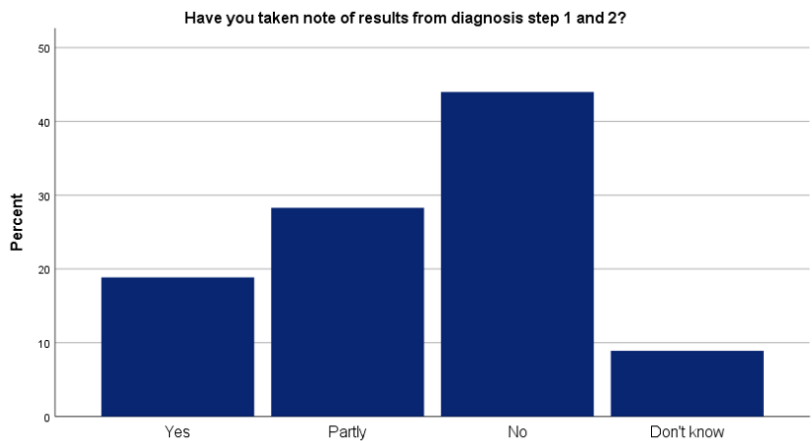


Figure 3.5 Teachers’ experience of using assessment data from the national diagnosis, “mapping”.

Collapsing the cases from the categories *yes* and *partly* shows that 47% used results from steps 1 and 2 ($N = 191$) regarding general academic literacy. As regards step 3 of the diagnosis, dealing with the students’ knowledge in the individual content subjects, the teachers had very little experience of performing the diagnosis: 7% answered *yes* and 9% answered *partly*, while 77% said *no* and 8% said *don’t know*. The results indicate that the teachers’ perceptions of the students’ literacy skills and competence is based on the teachers’ observations from class and their own assessments rather than the national diagnoses. It also reveals that the mandatory diagnosis procedure has not been fully implemented. Below, results are presented in relation to the teachers’ beliefs about the students’ language proficiency and the challenges associated with language requirements in different assignments.

Students’ language proficiency

Teacher beliefs about the students as learners have an impact on how they perceive the students’ abilities to progress and meet the knowledge requirements at different grade levels. This in turn supposedly influences how accommodations and language in instruction and assessment are dealt with, and to what extent the teachers believe language to be their responsibility as subject teachers.

Teacher beliefs about the students relate to three aspects in this study that are associated with the students’ language proficiency and thus ability to demonstrate their content knowledge and participate in activities in class. In the survey, the teachers were therefore asked how they perceive the level of

difficulty and the challenges associated with different written and oral assignment formats (Q20), the complexity of different language registers (Q21), and the students’ ability to use various verbal functions as found in the knowledge requirements (Q22). Results are presented in the following.

Four assignment formats were mentioned in the survey (Q20): *oral discussions in class, prepared oral presentations, short written assignments* and *longer written assignments (e.g., essays or reports)*. These were measured on a 5-point scale: *difficult, fairly difficult, neither easy nor difficult, fairly easy* and *easy* (see Table 3.13 below, where the assignments are organized in perceived order of difficulty, starting with the easiest at the top).

Table 3.13 *Perceived level of difficulty in different oral and written assignment formats (%)*

	Difficult	Fairly difficult	Neither easy nor difficult	Fairly easy	Easy
Short written assignments (N=190)	5	48	27	18	2
Prepared oral presentations (N=189)	20	47	21	12	-
Oral discussions in class (N=190)	31	46	15	7	1
Longer written assignments (N=189)	58	34	6	1	1

While *prepared oral presentations* was the only choice of the four that did not generate any responses for the *easy* option, *short written assignments* generated the least number of responses for the *difficult* option. Collapsing the *difficult* with the *fairly difficult* option, and the *fairly easy* with the *easy* option revealed that *short written assignments* were perceived as the least difficult of the four.

As regards the difficulty of different language registers, teachers’ perception of 6 different registers were investigated (Q21): *subject-specific concepts, subject-specific skills, subject-specific oral genre, subject-specific written genre, general academic language* and *everyday language*. These were measured on the same Likert scale as the previous question. See Table 3.14 below, again organized in perceived order of difficulty with the easiest at the top.

The percentages in the table below show that everyday language is the only language register that is perceived as fairly easy. Although all the others are perceived as almost equally difficult, it is interesting to note that subject-specific concepts are at the bottom, perceived as one of the most difficult. The subject-specific concepts are probably unknown to students of all languages as they start learning a new subject, but they represent individual vocabulary items and not more complex language production, as do all the other language features in the list. Subject-specific genre includes certain syntax as well as specific vocabulary and is perceived as fairly difficult, the written genre in particular.

Table 3.14 *Perceived level of difficulty in different language registers (%)*

	Difficult	Fairly difficult	Neither easy nor difficult	Fairly easy	Easy
Everyday language (N=191)	4	29	35	27	5
Subject specific oral genre (N=188)	31	53	12	3	1
General academic language (N=191)	34	45	18	3	-
Subject specific written genre (N=188)	43	43	12	3	-
Subject specific skills (N=191)	48	39	11	2	-
Subject specific concepts (N=191)	46	43	8	3	1

The subject-specific skills were exemplified in the survey as to *argue a case* or *draw conclusions*. These were examined in more detail in the next question: (Q22) *How often do you feel that your newly-arrived students have difficulty showing their content knowledge in the following activities?* Fifteen different action verbs or cognitive discourse functions (CDFs) were distinguished, using a 5-point scale: *always*, *often*, *sometimes*, *rarely* and *never*. Table 3.15 below presents a selection of action verbs that appear in all the subjects' syllabi, as exemplified previously in Table 1.9 (p. 54). These are presented in order of cognitive complexity in relation to lower- and higher-order thinking skills, with the lowest (i.e., least demanding) at the top, and not in order of perceived difficulty as previously (however, sometimes they overlap).

The order of appearance in Table 3.15 follows that of Table 1.9, which was adapted from the lower-versus higher-order thinking skills ranking scale by Anderson and Krathwohl (2001). Hence, it is not surprising that the first activity, *give an account*, representing the lowest-order thinking skill, displays the lowest figure in the first box (i.e., students do not have as much difficulty giving an account compared to the rest of the activities). Collapsing the first two—*always* and *often*—reveals that *analyze* is perceived as the most difficult activity.

Table 3.15 Perceived level of difficulty in different subject-specific skills (CDFs or function words) (%)

	Always	Often	Sometimes	Rarely	Never
Give an account (N=188) [<i>Återge*</i>]	4	35	42	18	2
Argue/give arguments (N=190)	14	57	25	4	-
Reason (N=190)	21	48	28	3	-
Describe (N=189)	10	42	40	7	1
Explain (N=190)	8	38	45	8	-
Draw conclusions (N=190)	20	48	26	6	-
Analyze (N=190)	25	55	16	3	-
Evaluate/make assessments (N=188)	15	54	26	4	1

*Here the Swedish word used in the survey is provided as the English translation is not as transparent as for the rest.

Following this presentation of data from the survey, the next section provides results from the interviews in relation to the same topic: teachers' experiences and perceptions of the students and challenges that they consider as they prepare to assess the students' content knowledge.

Perceived difficulty in different language skills and effects on grades

Language is key to communicating and learning in all subjects. As newcomers, the NAS start from scratch, learning an entirely new communication tool. The academic registers are considered to take twice as long to acquire as everyday language (Cummins, 2017) and must be taught in school, whereas everyday language can be acquired in other contexts, such as among friends or during leisure activities. In the survey, teachers perceive everyday language to be the easiest language register for NAS. In the interviews, however, several teachers point out that the NAS actually seem to be more well-versed in the language of schooling than in everyday language, probably due to a lack of exposure to Swedish outside of school or even the classroom. In the interviews, the teachers often use the term "varied" or "it varies" in reference to the significant differences in the NAS's conditions and abilities, as seen in the example from one of the social science teachers below.

- R What do you think about their [NAS] abilities and weaknesses, [and] how much they have advanced?
- T3 It varies a lot; some students have been in Sweden a year and are very driven and motivated, and they advance a lot, particularly in the language of schooling. They learn the more academic language, but not as much everyday language since they do not speak the language with friends or at home. Sometimes it is very surprising: "so you don't know that word?" Words you take for granted, but instead they know very difficult words or concepts from school subjects.

One of the physics teachers shared the same impression:

- T6 But there are all those Swedish words like slingshot, bouncy ball that [they] may not have heard of [...]. I know they are familiar with concepts like “the golden rule of mechanics” because that is part of what I am teaching, but those everyday words in Swedish, that’s what is missing and where there is confusion.

The teachers’ comments about the students’ lack of familiarity with everyday concepts due to poor exposure outside of school is slightly contradictory compared to the findings in the survey, where everyday language was perceived as the least difficult (cf. Table 3.14 above). This could imply that NAS lack the specific everyday vocabulary needed to have a frame of reference, as exemplified above in the interviews, but display knowledge of everyday “filler language.” Filler language is used as a strategy to paraphrase in the absence of disciplinary language, a phenomenon that has been noted among native language speakers as well, and is called a disciplinary inter-language (Olander & Ingerman, 2011).

The spread in the students’ proficiency becomes more obvious in the introductory program, posing challenges for the teachers, who sometimes meet illiterate students in need of individual solutions. One solution is to overlook the lack of writing ability and instead focus on oral performance:

- T4 Some of them are illiterate and can’t write, and they may totally flunk the test, but I may still notice that the student has followed all that we have discussed in class: “of course you will get a passing grade.” It doesn’t say anywhere in the knowledge requirements that they have to show their knowledge in written language.

The teacher above in language introduction looks for signs of content knowledge, even if it is expressed in an oral exchange in class instead of a written performance on a test. However, sometimes poor language skills prevent students from getting a passing grade:

- T11 I think we had 140 Fs [referring to the number of students with a failing grade of “F”] in the social sciences last year, so they don’t pass in the social sciences, and it is evident, if you can’t read and don’t have the reading comprehension, then you cannot get a grade.

The teacher above refers to Grades 7–9 in mainstream education. The teachers in the introductory program generally state that if the students attain higher grades they should be in mainstream education, not in language introduction. They mostly assign grades E to C. One of the physics teachers in Grades 10–12 points to the importance of language to be able to attain higher grades, and mentions giving one NAS an A, and several others having achieved B grades:

- T2 It varies a lot [the proficiency and ability of NAS]. The students who don’t have enough language skills they [...] rarely describe with words what it is they have done and calculated. Then again, there are other newly-arrived who have the language and know how to express themselves.
- R Does this affect your assessment in any way, to the advantage of the students who have enough language [...]?
- T2 Well, I guess so, if you take the concept “induced tension” for instance: to explain, calculate and use the formula is not the same as being able to explain what causes it [...] for that you need language.

In order to provide answers in relation to the second research question about how teachers' perceptions of the students' abilities affect their assessment beliefs and practices, teachers' perceptions are considered both in relation to specific challenges in relation to language, as above, and in relation to other background factors. One factor seems to relate to a different perception of what it means 'to know something' among the NAS, as outlined below.

Age, varying scholastic background, perception of knowledge and students' progress

Based on their experiences of the students in class, the interviewed teachers mention different aspects that seem to have an impact on the NAS's academic achievement. The teachers refer to the students' previous schooling and time of arrival. The students' age at the time of arrival, implying late time of onset to acquire the language of instruction, is often mentioned as a critical factor for learning subject content in a new language (Bergendorff, 2014; Cummins, 2017; SOU 2017:54). One of the teachers comments that native Swedish students have been able to practice the required skills since Grade 1 (e.g., critical thinking and argumentative skills), but the NAS have rarely practiced those skills at all, which will prevent them from reaching higher performance levels in assessment, where argumentative language skills are needed. However, the fact that they are older means that they are at a more advanced intellectual level and can progress faster, according to the same teacher. As regards the NAS's progression to more advanced levels of knowledge and thus their ability to attain higher grades, the teachers point to the importance of previous schooling and literacy development in their home countries. Several of the teachers, as in the example below, mention that school background and home situation affect how fast the students progress:

- T3 It is not so much about time, but previous experiences; those who have a good educational background are much better equipped. [If they don't have a written language when they come] then it is a matter of getting them to pass the course so they can apply to vocational training in upper secondary.

Another teacher refers to the challenges associated with the heterogeneity of the groups in the introductory program:

- T4 [There is a] very big spectrum in the groups; one group had a spread in language proficiency 2, 3 and 5 [their own levels] [which] means that there was a lot to take into account, and [it had to be] very individualized. It has been a learning experience.

A general view expressed by the teachers in the interviews concerns what they identify as a different perception of knowledge among the NAS. Whereas Swedish syllabi advocate critical thinking (the ability to apply knowledge and reason), the teachers note that the NAS focus on what can be read in the course book and what the texts say explicitly. The NAS are very skilled at memorizing and believe reading many pages will grant them good grades, according to the teachers. When students are asked to reflect, draw conclusions and apply their knowledge by reading beyond the lines, the teachers note that a typical response from the students is "but the book didn't say so" or "but it says so on page 78 in the book." The students seem unused to the idea of drawing their own conclusions or presenting alternative views:

T5 I have noticed in my students that their previous schooling was so much more factual—you need to know facts and you need to be fast in stating and explaining ALL you know—while I come up with all the “Why?” [...] It is really difficult for them.

R In what way is it difficult?

T5 Because they don’t think; well, they are more factually oriented: “Are you supposed to think freely, see consequences, look at differences? [...] Can you have a different perspective than what the teacher is presenting?”

The same teacher comments on the difficulty in explaining the qualitative differences between factual recall and critical thinking to the students:

T5 The students don’t understand why they don’t get a passing grade, they don’t understand the Swedish education system. “I answered all the questions on the test, why can’t I get a grade, why can’t I move on?”

The interviewed teachers mention challenges they are facing in trying to get the NAS to understand why factual recall is not enough to attain a higher grade. They are wrestling with how to make NAS understand the difference between quantity and quality at times. It can be very frustrating for the teachers, since they know that the reward for the students’ hard work takes time if it is ever to pay off the way the students are hoping:

T11 They know that they were good in their homeland, and then they come to us and they want the highest grade, and there is a discrepancy that we have to work with all the time. We can’t explain how it works in Sweden, that you must achieve these things. Speaking of concepts, how do you explain “[ability] to reason” to someone who doesn’t understand? It is really difficult [...] they think they know things because they have answered questions.

The teachers share the experience that it is difficult to explain why factual recall is not enough or to make the students understand the qualitative nuances in the course goals that require higher-order thinking skills or more elaborate language in order to attain higher grade levels. How do you get through to the students when they lack understanding of the phenomenon (i.e., drawing your own conclusions, applying knowledge and “thinking outside of the box”) as well as the language? This also relates to the next issue brought up by the teachers at the interviews: the students’ unwillingness to expose their challenges or reveal that they are struggling:

T5 Those who don’t have that much language background, they are not so willing to ask, they don’t want to show their lack of knowledge.

This makes it more difficult to know how to help the NAS and make accommodations in the assessment situations, according to the teachers. At the same time, teachers express appreciation for the motivation and eagerness certain students show to learn. Sometimes they mention that the NAS show greater drive than many native Swedish students. The interviewed teachers demonstrate great commitment to helping the students and the almost unsurmountable task at times of bridging both the performance and cultural gaps, which is shown both in gestures at the interviews and the use of words such as “frustrating.”

When talking about factors that affect students’ ability to perform in assessment, different environmental experiences are mentioned. Besides the ones stated above, lack of ordinary everyday experiences is mentioned, as outlined below.

Lack of cultural cues and everyday experiences

Lack of cultural cues and a frame of reference to relate new knowledge to previous experiences is mentioned as an issue in relation to equitable assessment, according to the literature (Gipps, 1994). This is something that the NAS obviously lack in comparison to native Swedish students, according to several of the teachers:

- T1 When you explain something, if I have a group with only Swedish native speakers, then you realize how much previous knowledge they have. I have a group of students, vocational students, they have some other issues, but when I explain something, they get it immediately. That never happens with a group of immigrants, regardless of the program. You always have to think, no, they don't have the background knowledge.
- R Can you give an example of what this could be?
- T1 Everything that has to do with traditions, Swedish traditions [...] it is basically the unspoken, the things we never talk about when we meet Swedes. We know the context [...] the immigrants don't [...] and you can't expect that from them, THAT is the big difference.

Another SS teacher in the introductory program mentions how she sees the students' experiences from their countries as a resource that is invaluable for the students to retain knowledge. She uses the metaphor "hangers on the wall":

- T5 They get to do a lot of comparisons with their own country [...] I used to think that they have hangers on the wall; if there are no hangers, the jacket falls flat to the floor, it is the knowledge falling flat.

Using the students' previous experiences is recommended for reasons of equity and inclusion (Bergendorff, 2014). Moreover, the cognitive and mental processes are closely associated with everyday experiences from life; if these references are missing, it complicates learning subject content, as mentioned here in physics:

- T7 You need language to discuss, and things like "catch up" and such expressions can be really difficult. Now there are phones, so you can run a Swedish-Persian translation to show what it says, and that way you can go far, but these discussions get more complicated when you deal with everyday experiences you have from childhood: how to bike and swim, do ordinary things that not all may have done, for different reasons.
- R So it relates both to experience and the phenomenon?
- T7 Yes, that's what I feel.

The teacher further notes that lack of a frame of reference is not only a concern with the NAS but is also true for certain groups of native Swedish students, thereby hinting at different learning conditions caused by socioeconomic differences:

- T7 Even if you are from Sweden you may have poor language [because] you never discuss those matters at home [...] Maybe you have other discussions around the dinner table [...] hunting or farming or something else, I don't know, maybe some trip abroad, or TV show, there is so much you can talk about.
- R But not physics?
- T7 Not physics.

One of the teachers who has experience in teaching both in the social sciences and language expresses the same thing and draws the conclusion that all students need teachers who see themselves as language teachers. It is important to provide all students with equal opportunities.

In summary, the teachers state that issues in relation to poor language are key, since this means that the NAS are inhibited from expressing themselves to show their content knowledge. Challenges are associated with the lack of exposure to everyday concepts that are usually taken for granted, according to the teachers at the interviews. In the survey, the teachers reported that subject-specific genre—both oral and written, but particularly written—are difficult for the NAS. It is further noted that the students generally have difficulty understanding the meaning of higher-order thinking skills, since they come from contexts with a focus on factual recall, according to the teachers. The teachers rarely see the NAS attaining above a C grade, although it does happen. The conditions vary largely due to differences in scholastic background and the heterogeneity of the group. Taken together, these are aspects in relation to the students that may have an effect on teachers' assessment beliefs and practices in the subjects.

Before delving into assessment, the key area of interest, a second aspect mentioned as affecting teachers' assessment beliefs in the second research question is outlined below; it relates to language: How are the teachers' self-reported assessment practices affected by their *beliefs about language policy in education* (i.e., national curricula, syllabi, translanguaging practices and local organization, where the latter refers to collegial collaboration and responsibility for language development).

Teachers' experiences and beliefs about language in education

The second research question implies that teachers' assessment practices are affected by their beliefs about language in education (as noted above) and how language is dealt with in the local organization. An important question is thus *what* language(s), and *how* and *when* is it processed and taught, if indeed this is considered part of assessment? This will be investigated in the next section. If students are not given the opportunity to learn the language that is required for the assessment, language becomes a potential construct-irrelevant variance, which implies a threat to the validity of the assessment (Crooks et al., 1996; Gipps, 1994; Stobart, 2005). Language in education thus relates to language skills in relation to the subject and its content and what is required of different assessment formats in terms of language. It also relates to access opportunities as regards which languages are allowed or expected to be used. In relation to access opportunities, questions arise as to who is believed to be responsible for providing students with the language needed. Below, teachers' beliefs about language in education, and thus in relation to assessment, are presented in order of appearance in the research question, starting with curriculum, syllabi and activities in class.

Disciplinary language taught and used in class

In the survey, the teachers were asked how they prepare students for assessment and how frequently they do certain things in class as regards certain language features and registers (Q32); see Table 3.16 below. The activities are presented in order of frequency, with the most common at the top.

Table 3.16 How teachers prepare the students for assessment tasks (%)

	Always	Often	Sometimes	Rarely	Never
Looking at subject-specific concepts (<i>N</i> = 190)	76	18	5	1	1
Presenting useful sentence structures (<i>N</i> = 188)	27	39	25	6	3
Providing written examples and models of subject-specific texts (<i>N</i> = 190)	22	30	33	10	5
Writing texts together in class (<i>N</i> = 188)	13	27	40	15	4
Providing oral examples (e.g., how to make an argumentative speech) (<i>N</i> = 188)	14	25	38	14	9
Looking at the structure and linguistic features of subject-specific texts (<i>N</i> = 189)	15	27	25	26	7

In the first question about subject-specific concepts, 76% (*n* = 144) of the teachers reported *always* spending time on them and 18% (*n* = 34) reported *often* spending time on them. A majority also seem to present useful sentence structures, which 27% (*n* = 51) and 39% (*n* = 73) said they do *always* or *often*, respectively; this represents more than half of the teachers. Only 9% (*n* = 17) said they *never* or *rarely* do it. A little over 50% (*n* = 99) also state that they provide written examples and model texts, collapsing the cases in the first and second columns. The third aspect relating to the study of texts—examining their structure and linguistic features together in class—displays the highest negative result, with 32% (*n* = 62) answering *rarely* or *never*.

The results indicate that language instruction in the subject content classes is generally about subject-specific concepts and so called “language chunks,” as in useful sentences. When it comes to the development of subject literacy, including the writing of coherent texts, studying the character of longer disciplinary texts or providing oral argument examples, the teachers’ responses indicate that this is not frequently dealt with in class.

When comparing teachers across disciplines, a certain difference can be observed: 82% (*n* = 46) of the SS teachers always provide concepts, compared to 68% (*n* = 53) of NS teachers. SS teachers provide more oral examples (38%, *n* = 21) than NS teachers (24%, *N* = 19) when collapsing cases for *always* and *often*. The reverse can be observed for presenting useful phrases and providing written examples; collapsing cases for *always* and *often* shows that 68% (*n* = 53) of the NS teachers present useful phrases, compared to 53% (*n* = 29) of their SS counterparts. The NS teachers provide written examples slightly more frequently (45%, *n* = 35) compared to the SS teachers (39%, *n* = 22). The teachers with a dual combination (subject and language) display higher frequencies across all activities (cf. p. 173, *Comparison across teacher groups*).

A lot of focus is on concepts and the teaching of key terminology, but sometimes verbal skills are included as expressed in the knowledge requirements. A couple of teachers mention how they try to take

this a step further by teaching general academic language and teaching linking words, for instance, like this teacher in social studies, who says he speaks for his colleagues as well:

- T13 We are trying to work more on language development, and it is not only because of the newly-arrived, but for all students, to help them express themselves and so forth. We read a lot and we are trying to work on concepts, so they learn to understand concepts both when reading and for them to use in texts, but also, well, simply words to use when you explain things. We have worked with linking words, for instance, especially in ninth grade, words they can use to develop their answers, so we are trying to work with language development for all students.

The teacher notes that he is trying to model reasoning skills. Like most of the other teachers interviewed, he acknowledges that it is challenging, but refers to the knowledge requirements in the syllabi:

- T13 When I work with language development, I try to make them think about how A will lead to B, which will lead to C, but can also lead to D, or something like that; trying to model the reasoning process, but it is difficult. I don't have that much experience either, so it is tricky, but this is what is required in the knowledge requirements.

One of the other teachers in language introduction says it is difficult to model and teach higher-order thinking skills due to the limited language of the students and hence the basic level of the instruction:

- T4 I did write down general descriptions for the different levels, with the intention of translating this into their mother tongues, but it never happened, unfortunately; but to illustrate... well, again it is a matter of their language proficiency [...]. When instruction is at a very basic level without discussion and problematizing in the lectures, then it is harder, in a group that is less proficient in the language.

The teachers who have NAS integrated into regular classes note that all students profit from a focus on language. Most teachers have a strategy as to how to integrate this into their instruction, but not all:

- R Do you talk about these things [language] in class?
T7 No, we should do it more; I stink at that, I am probably a little ignorant [...] all special ed teachers and such, they say "make wordlists and then you work on it for a while" and then you do it for a while, and then, ouch, I haven't done a list in a long time, you know. It would be good to make good lists with all the concepts that are included in the course, and what they mean.
R Do the students make lists themselves?
T7 Some do, but we should be more purposeful with this.

The teacher, who teaches physics in the natural science program in Grades 10–12, shows an awareness of the importance of language and acknowledges a need to be more purposeful than he is currently. He, as well as the other physics teacher in upper secondary, notes that the subject-specific language is used in class, suggesting that the students should be able to pick it up; they are referring to the students' personal responsibility. The students should learn the relevant language, since these are typical expressions that are used in class, according to them, thereby signaling that language is acquired incidentally. When it comes to the "other words" that are needed to formulate answers on a test, for instance, one of the teachers explains that he tries to think ahead so he can adjust the test questions to better fit the students, but notes that there is a risk that this may be neglected until the need is revealed in the assessment situation:

- R Before the test, have you gone through language or concepts because you notice that they don't understand certain concepts, or do you need to take time [during the test] to explain?
- T2 I often feel that concepts are easier to understand, the terminology you find in the course book, that is explained there, they learn THAT. The other words, that may show up in a test, that [is something] I might not think about until they are taking the actual test.
- R When they ask, you mean?
- T2 I try to think about it before the test, when they ask in class, because then I realize how I may put the question differently, but they can ask questions when they take the test, because I can't, I might not have time to consider everything [beforehand].

Some of the teachers explain how they try to teach the students to “think outside the box” and use their own words, not just copy and paste, by discussing relationships of cause and effect together in class or by modelling examples of what it may look like to reason:

- T3 After a test I used to show them, “this is an example of a good answer to this question,” “this is how you can answer at E level, or C level and so forth,” for them to see. Before a test too, if it is possible, I give examples all the time [...] This is how I would answer the question, now it is your turn. Lots of examples, modelling. My students have learned this strategy and go to other teachers and tell them “we need to know what it looks like, we want examples.”

The teacher in the example above refers to the teaching and learning cycle, TLC (cf. Part 1, p. 25 and Part 3, p.100), as do several of the teachers who also teach SSL, which means going from context exploration to text exploration, on to joint construction and, finally, individual application. Some teachers, on the other hand, are concerned about the students' tendency to copy if they are presented with a model text, in some cases commenting that this may lead to plagiarism or pure memorization, thus leading to factual recall instead of reflective thinking. Another teacher shows awareness of the copy and paste strategy but seems to be able to cope with it, although it takes time.

- T3 They learn that it is good [an example], and they learn how to use them in a good way, not only copy, but how to use the examples, and this takes a while, to teach them this, and how to use your own words, and not only copy and paste. These are things Swedish students learn from Grade 1, and these students need to learn it simultaneously with other things, since they come maybe in Grade 5 or 6. There is so much they need to learn.

When language is taught

When asked about how and when language is taught in class, the teachers have different approaches. Most teachers seem to do like the teacher above, following up after assignments or using test results for formative feedback. Some teachers state that they introduce every new theme or class with an introduction that focuses on concepts and skills, which again resembles the TLC approach (see above):

- T12 I used to tell in the introduction of every class which particular skills we were going to practice that class; if we are working on the ability to reason, I write that on the board and “this is what we are working on today,” those are the skills.

- T4 Regardless of which group in language introduction (but I do it in regular upper secondary classes more), I start with a presentation of key concepts. I write them [on the board], and in language introduction I often start with “today I am going to talk about transnational companies, multinational companies, globalization, etc. You have 15 minutes to Google, find out the meaning and write it down, and then I want for us to share what that means.” Then I write down the explanation on the board. After this half hour, or whatever it may take, we look at them together, and it becomes student-centered; we practice the concepts while I am also trying to provide context.

Others choose to make sure the students have picked up some concepts while listening to the teacher and discussions in class before they go any deeper into language, believing in scaffolding and letting the students do some thinking for themselves before presenting more language “in the middle” of a thematic unit:

- T5 I usually do it in the middle, not in the beginning, because then they don’t have all the concepts and we still do a lot of talking, and they get the foundation [...]. It is also part of having them think freely, that it is not a matter of trying to learn things by heart. Instead, I must work on my knowledge, and they need that, or else they will only read what it says and copy. They tend to do that a lot, memorizing.

All teachers seem aware of a need to provide the students with language and many do, both subject concepts and subject-specific skills as found in the syllabi; but do they let the students use other linguistic repertoire, such as in the students’ first language or something else to bridge the language gap, or is the focus mainly on Swedish?

Language use in class and translanguaging

When asked if the students can use their first language, if they must use Swedish, or if they can mix as long as the students make themselves understood, the teachers express varying experiences and practices. Some teachers are very clear about the goal for all students to learn Swedish and feel that the fastest way to integrate is to avoid using their first language as much as possible. Others state that *if* it is their right to use the language they are most comfortable using, they should, but comment that it makes it more difficult for them as the teacher to control what is learned, especially in assessment situations. The word “if” indicates that uncertainty prevails as regards what languages are allowed and advocated in class. Two of the teachers in the natural science program are either skeptical or note that the use of the students’ strongest language has not seemed to be an option:

- T7 That’s the kind of thing I don’t think about much; I should do it more, I am sure it is great, but I am so busy doing other things, there is grading and stuff.
- T2 I am a bit skeptical. I want them to speak Swedish during my class and use the Swedish language.

The same teacher comments that in a class with a lot of Arabic-speaking students, their use of Arabic would exclude other students and the teacher, who do not know the language:

- T2 We explain to them how important it is that in this school we speak Swedish and “it will only affect yourselves if you don’t learn Swedish properly when you finish school” [...]. Mixing many languages at the same time might work in the introductory program, but once you make it to upper secondary, I don’t believe it is right. Then, you need to know at least as much Swedish as the grade you have been awarded represents.

One of the teachers in the introductory program establishes that the goal is to use Swedish, which the students should do as well:

- T1 I speak Swedish, but sometimes they use a number of languages between them, but I used to say that I cannot follow then.
 R Are they allowed to use other languages?
 T1 No.
 R ...in the classroom?
 T1 Well, if they can... they do it anyways, but I try to tell them that we stick to this language, that is enough, this is the language you have to know in this country.
 R You mean Swedish?
 T1 Yes, that is Swedish.

Sometimes the use of the students’ L1 is seen as an accommodation when students are new, but the teacher in the example below maintains that it only applies during the introductory phase and that it is important to be integrated in Swedish as soon as possible:

- T10 They let go of their first language the more they learn [...] and those who are strategic [...] try to get friends with a different first language than themselves. The students with the slowest progression are often those who stay with their peers who share the same language.

One of the teachers mentions how English-medium instruction is used as a type of accommodation and solution to bridge the language barrier:

- T6 In the beginning, I wasn’t able to make that many adjustments, since you have to teach, and they [the students] didn’t know much English. One knew a little English, so one of the accommodations was that I used English as a medium of instruction [and] taught in Swedish AND English simultaneously.
 R Did you go back and forth between languages?
 T6 Yes, I used both Swedish and English for the whole class. I actually repeated in English what I had said. I did that in another class as well [where there was a girl] [...] who knew English fluently, [...] so I taught chemistry in both English and Swedish.
 R Did you receive any reactions from the Swedish-speaking students?
 T6 Absolutely, in the beginning, but then it faded, or they thought “wow, you can speak English?” [...] They only [...] reacted positively, not negatively; they thought it was fun.

Using different languages in class can be described as a teaching strategy or an accommodation when language is insufficient. Questions of fairness may be posed, especially when different conditions apply depending on the language and where the students come from (cf. Kaya, 2017), something several teachers are aware of:

- T11 There is no e-lexicon, no audio reading. They are there and what can you do, they have Google Translate, and that affects the results too [...] he [a boy from Thailand] used his cell phone to translate.

Another teacher talks about how it is unfair because the Arabic-speaking students have a lot of material and support, whereas a boy who came from Albania didn't have anything in his language:

T3 You can't deny the other students the material just because there is none in his language.

Once again, the teachers' practices in class seem to vary depending on school level, resources and the teachers' background and subject profile. When there is access to study counselors or translators, they assist, particularly during early introduction of students; however, it seems more common that teachers must solve language issues themselves. According to the Swedish School Ordinance, NAS are supposed to be granted study counselling in their mother tongue (SFS 2011:185, Ch 5, §4). Below, some findings are presented in relation to how teachers perceive of collaboration with colleagues and responsibility for the students' language development.

Organization and collaboration—who is responsible for language?

The context seems to influence how the responsibility for the students' language acquisition is perceived: If the student is in the introductory program or integrated into a regular class, if it is Grades 7 to 9, or 10 to 12 or if the student is a recent newcomer and within the first months of attending mainstream education. The school level matters, but language instruction and accommodations also depend on the organization and resources of the school. The teacher's subject also seems to influence the language approach.

When the teachers were asked if they collaborated with a colleague when planning instruction and/or assessment of NAS (Q36), 19% ($n = 36$) reported always doing so and 35% ($n = 65$) reported doing so often, while 29% ($n = 54$) reported doing so sometimes; this indicates that teachers collaborate quite a lot. When asked who they collaborate with, the teachers could indicate as many of the following as they wished: *study counselor*, *mother tongue instructor*, *SSL teacher*, *special ed teacher* and *subject matter colleague*; the subject matter colleague was the most frequent collaboration partner. Whereas 84% ($n = 119$) stated that they collaborated with a subject matter colleague, only 21% ($n = 30$) reported collaborating with a study counselor and 28% ($n = 39$) with a Swedish language colleague (e.g., an SSL teacher); just 6% ($n = 8$) stated that they collaborated with a mother tongue instructor. A couple of other questions in the survey asked who the teachers collaborated with in relation to different aspects of education. Here, the response rate was low, with a lot of missing values due to nonresponse, which might be explained by whether the teachers perceived that the question was applicable to their organization. The lack of response could also be explained by the position of these items toward the end of an extensive questionnaire. Regardless, the table below offers some data from the smaller sample, (frequencies in valid percent based on the number of respondents). The question (Q37) asked if, and if so, with whom, the teachers collaborate; they could respond *yes* or *no* in relation to each colleague and feature presented to the left.

Table 3.17 *Who the teachers collaborate with in different language-related tasks (%)*

	Study counselor	Mother tongue instructor	SSL teacher/ Swedish language teacher	Special ed teacher	Subject matter colleague
Ideas on how to work with subject-specific language in class (<i>N</i> = 134)	11	6	38	19	75
Ideas on how to work with general academic language in class (<i>N</i> = 110)	7	4	32	26	71
The colleague helps review the language used in instructions, questions and assignments (<i>N</i> = 97)	19	11	36	5	58
The colleague helps the students with vocabulary lists in their language (<i>N</i> = 110)	54	27	29	4	12
The colleague helps the students read and/or write texts (<i>N</i> = 125)	50	26	34	11	20
The colleague helps assess language in assignments used for assessment (<i>N</i> = 96)	21	14	33	3	53

It is striking that the teachers prefer to collaborate with subject matter colleagues (at least in tasks that relate to classroom activities and teaching material), as seen to the right in Table 3.17. The response rate dropped in relation to language in instructions and assignments (the second feature), which may suggest that this is not a common procedure. It seems that the teachers believe that study counselors are best suited to help the students with their language development. In relation to assessment, the last feature, the response rate dropped again, but if someone is involved, it is the subject matter colleague. The SSL/Swedish language colleague seems to be perceived as a support in relation to all features by one-third of respondents on average, making the language teachers appear to be generalists who are asked about various matters.

In the interviews, the teachers reported that during the early phase, when the student is first introduced into a regular class, a study counselor or mother tongue instructor is often present in class as a translator. Later, it seems an SSL teacher carries more of the responsibility in helping the students, but it varies as to whether or not the teacher follows the NAS to the subject content classes. Most of the time, the SSL teaching seems to be separate, according to teacher reports. When it comes to the teachers' sense of their responsibilities as regards disciplinary language, all the interviewed teachers in the study recognize a need to make the students aware of key concepts, not just the NAS, which aligns with the results in the survey (cf. Table 3:16).

- T9 We work a lot with concepts, I try to include a lot of more conceptual lists and vocabulary lists [...] that we work with before we start a new theme.
- R Is this something all students do?
- T9 Yes, everyone does it, I think it is important for everyone [...] there are many students who need to work on this.

Teachers refer to SSL colleagues, but they also express an apparent lack of study counselors, especially now, a couple of years after the big surge, when it seems that the organization has been slimmed. One of the teachers in the introductory program mentions that helping the students with language is evidently part of the job, although he cannot do it all. Instead, he refers to language instruction that is separate but offered by the school:

- T1 That is separate. A retired principal and an older former politician [...] come and help with special community information on Tuesdays after school [...] so that is a way to boost it.
- R But do they teach language?
- T1 Yes, language and its content.
[...]
- T1 Last time they were teaching the parts of the body.
- R Is that mandatory or voluntary [for the students]?
- T1 It is voluntary, I guess.

The students' language development needs all possible boosting, as the teacher noted in the quote above, but a general question after listening to what the teachers have to say concerns the varied and unequal practices and resources at different schools, and hence what this entails for equity in assessment.

In summary, all teachers see limited language as an obstacle for assessment that typically puts the NAS at a disadvantage. The teachers report varying beliefs as regards which language should be used and whether the mixing of languages can be considered a possible accommodation. All teachers see Swedish as the norm and the target language; becoming proficient in the subject skills in Swedish is part of the intended learning outcome (ILO). The teachers express slightly different approaches to language in class, and perceptions of whose responsibility it is to make the students "subject literate." Below, questions investigating teachers' assessment beliefs and practices are outlined.

Teachers' assessment beliefs and practices across subjects and schools

When investigating teachers' assessment beliefs and practices, certain contextual factors have to be considered, such as local policy and organization. This, together with the teachers' individual interpretation of intended learning outcomes in the subjects, views of learning and of the students, shapes how assessment is performed. Below, the teachers' assessment beliefs and assessment practices are discussed in relation to *what* it is they believe should be assessed and *how* this should be done depending on the purpose of assessment and the prevailing conditions, which are shaped both by the needs of the students and local policy. The presentation of teachers' assessment beliefs starts with what they believe the students need to know in their subjects.

The “what” of assessment and the intended learning outcomes

Teacher beliefs about assessment overlap with what the teachers believe the students need to know and what they in turn need to teach; in other words, they are closely connected with the objectives expressed in the knowledge requirements and core content in the syllabi. The ILOs are equivalent to what the students should know at the end of a course. The first step in providing well-planned instruction, according to constructive alignment, is to start with the goal and ensure that learning activities in class and assessment align with the intended curriculum.

In the survey, the teachers were asked how important they thought it was for the students to be able to use different language registers to show their content knowledge: (Q28) *How important are the following for assessment of students' content knowledge in your subject(s), on a scale from very important to not important?* Four different language registers were included: *subject-specific concepts; subject-specific skills (e.g., to argue, draw conclusions); subject-specific genre (e.g., lab reports, historical chronicles); and general academic language*. The teachers could respond in relation to each of the language registers and thus state that all four were very important, if they chose. The language registers are presented in perceived order of importance, with the most important at the top.

Table 3.18 What teachers believe is important for students to know for assessment in their subject (%)

<i>That the student can use...</i>	Very important	Fairly important	Not important
Subject-specific skills (e.g., to argue, draw conclusions) (N = 191)	72	28	-
Subject-specific concepts (N = 189)	59	39	2
General academic language (N = 191)	31	57	13
Subject-specific genre (e.g., lab reports, historical chronicles) (N = 189)	27	55	19

Being able to use subject-specific skills was perceived as particularly important. This can be compared with the frequency for subject-specific concepts, which actually was lower, and subject-specific genre, which 27% ($n = 51$) called *very important* and 19% ($n = 35$) went so far as to call *not important*.

In relation to language, the teachers were asked if they give corrective feedback on the students' language in different situations, both in class and for assessments: (Q30) *How often do you do the following with your newly-arrived students?* Results are presented in order of frequency in Table 3.19.

Table 3.19 Frequency of teachers' corrective feedback on students' language (%)

	Always	Often	Sometimes	Rarely	Never
Give feedback/correct students' language on written assignments (<i>N</i> = 189)	21	31	29	13	6
Give feedback/correct students' language in class (<i>N</i> = 190)	11	31	33	17	9
Correct the students' choice of vocabulary on assessment tasks/tests (<i>N</i> = 188)	5	23	38	23	10
Correct the students' language (e.g., spelling and grammar) on assessment tasks/tests (<i>N</i> = 190)	6	14	27	25	27

As seen in Table 3.19, teachers display varying practices in terms of their perceived frequency. Written assignments seem to generate the most corrective feedback and the least is given for tests or assessment assignments, according to the teachers. This may be explained by the teachers' intention of separating language and content, thus preventing the students from getting the idea that language is being assessed. However, the teachers do acknowledge correcting language to various degrees. Depending on stance, this either confirms that all teachers can be language teachers in a sense, or else indicates that for this group of students, or maybe for all, the differing amount of linguistic feedback may as a matter of fact represent inequitable treatment and be unfair in relation to the compensatory mission (cf. Andersson et al., 2015) of bridging the gap for the students. Views as to which language the students should be encouraged and equipped to use and whose job it is to develop the students' language depend on ideology and beliefs.

In the interviews, the teachers were not asked explicitly about ILOs; when talking about assessment, however, it becomes clear what they perceive of as learning objectives. Teachers are unanimous about the focus of attention and the aim of assessment:

- T5 You look at the knowledge requirements, whether they [the students] have very good knowledge [...]. You start from the course plan: *has very good knowledge*; well, then they have grasped a lot of the core content and they have been able to develop [this] [...]. The tricky part is when you develop the tests to be able to make questions that will generate answers that will show if they have reached a higher level.

Challenges in relation to language and higher-order skills

According to all the teachers, in order to earn a higher grade, above an E, the students need to be able to reason and discuss. Even though Swedish syllabi require the ability to apply knowledge through reasoning to some degree at all levels, even for the lowest passing grade (E), teachers seem to agree that for an E it may be sufficient to not use as much language and show more conceptual knowledge or factual recall. One of the NS teacher talks about the impact of language on the subject and compares the social sciences and the natural sciences, claiming that the latter is easier for NAS since it requires less linguistic elaboration. However, he states that assessment in his subject involves language too, which is

what hinders NAS from attaining higher grades. He mentions two NAS who were ambitious and had good previous schooling, yet received an E and a D:

- T6 They received an E and a D, and their challenge is language related. Assignments dealing with problem solving in math are difficult, to reason in several steps in the natural sciences, it works, but you need to use concepts and it needs to be... you are actually assessing Swedish too. It works though, since the subject-specific language in the natural sciences actually is pretty concise; you don't need to write long texts about things like in the social sciences and language classes, where you have to elaborate. In the natural sciences, it is better if you don't.

Or as this SS teacher in the introductory program states:

- T1 I try to make them see how they develop [by thinking outside the box], how to reflect more. I am after the critical thinking, not just learn fact and fact. [...] What I assess is... if they only do factual recall, well then it corresponds to an E or D, but then, the more they think outside the box, and share their own comments, well then it is a higher grade. That is basically how I go about it.

When it comes to the students' ability to perform and how teachers can make valid interpretations of the results on tests or assignments used for assessment of subject content knowledge and skills, various challenges are mentioned, most of which are language related. This has to do with the use of cognitive discourse functions and being able to assess the more advanced skills, as expressed by the teacher above (T6) and this teacher in language introduction:

- T8 The difficult part in assessment is how to deal with the analytical [skills] and discussion, since there are facts and discussion, but there is no... analysis, there is no depth, no creation of new material. We work toward something that must be taught, depending on whether there is a test or assignment: factual recall, either by heart, or it if is a test or work in a group, with help from some material. The biggest challenge [in language introduction] is to create quality in the teaching.
- R How did you tackle that?
- T8 Well, with tasks that require some development, where they need to think, compare. There might be two known variables, and I want them to create a third, give an opinion. A and B are known, and they must come up with C, based on A and B.

Limited language and thus inability to show higher-order thinking skills are identified as inhibiting the students' ability to reach the more advanced levels. Teachers also mention limited understanding of what to do and what is required in a question or a lab situation, which can relate to cultural cues together with an inability to produce meaningful language:

- R What do you think is most difficult when it comes to assessment?
- T6 It could be... hard to say..., when they make language mistakes [...] if they haven't learned what it is, they can mean one thing and then write something else. So, to really assess knowledge and not language is difficult. But [...] all these [students] [...] have had this opportunity to take the written test, have me go over it with them while they were taking it, to check that they have understood the questions, and afterward go over it again and get the opportunity to show in oral supplementary tasks if they feel they weren't able to express themselves [in writing].

The teacher mentions the use of several accommodations to help the students both during the test, when the teacher offers help explaining, and after, when the students are given the opportunity to make clarifications. Teachers also mention challenges in practical lab situations and the ability to plan:

- T6 We had big issues. They manage the theoretical pretty well after a while, [...] comprehension, while they don't have a clue what to do in practical tasks, like labs in chemistry [...]. Things many find pretty easy in Sweden, but that was like, hey, what is this, but it worked out. But you could see differences in... okay, they were able to reason in multiple steps, express width and depth in theoretical reasoning, but when they were doing a lab, they didn't know how to plan.

Teachers also identify students' challenges in decoding questions and understanding implicit requirements in the subject, which adds to the already poor linguistic resources for explaining the process:

- T2 It doesn't say to draw a figure, but in order to solve the question I think you need to draw a figure too [...] sometimes they look up a formula and use it, and sometimes it works, but the question is, did they understand? You can't tell, because some have poor explanations, but they may have calculated it correctly [or] one student may have drawn a bad figure, and then the whole question is misinterpreted because of that.

Challenges in relation to accountability and students' memorization strategy

Several teachers express issues in assessment related to the strategy used by NAS of memorizing and copying, which was mentioned previously in relation to teachers' perceptions of the students. One teacher in the natural sciences mentions assessment of assignments they have done at home and the fear of plagiarism:

- R Are you afraid they may copy the text then?
T2 Well, especially the language, [...] when you see that the language is perfect you become suspicious, especially among the newly arrived, who don't master Swedish, and then it is important to point out that I do not correct spelling mistakes, or vocabulary mistakes, I look at the whole picture.

This reveals a kind of contradiction teachers are facing in assessment with NAS. On the one hand, certain teachers acknowledge the need to model language for the students to learn, but on the other hand, there is the fear of giving the students credit for language that is not their own, but the result of a copy and paste strategy, without them having internalized the language. Ultimately, it is a matter of students showing their understanding, which can be done by giving the students oral follow-up questions, according to the teachers. As regards copying, one of the teachers in language introduction acknowledges the possible conflict between the students' learning strategy and accountability in assessment. He says that he gives them the same questions in class that will show up on the test for them to practice:

- T8 I tell them "I want you to learn this by heart", [but] there is a problem, legal certainty, is this okay to do? I think it is, if my purpose is to get them ready for a national program [...] I know they will learn instrumentally since their language is so poor, they will memorize.

It is often maintained that expectations must be kept at a high level for NAS, and there should be no lowering of the bar. One of the teachers says that even if the NAS feel they cannot write long texts, you need to push them, even if it takes time:

T9 You have high expectations for them, but at the same time you cannot stress them out either; it has to take the time they need to learn.

To summarize what teachers believe it is they assess, they all refer to the subject-specific skills and the progression expressed in the knowledge requirements, both in terms of the development of higher-order thinking skills and in relation to the qualitative descriptors (e.g., *in basic terms* vs. *in detail and simple* vs. *well-grounded and balanced*). Some teachers comment that it is difficult to decide what simple versus well-grounded implies in reality and what should be required of the students in this context: “For a higher grade, does it have to be perfect?” as one teacher asked.

Teachers mention that they start planning their assessment by breaking down the core content, and when they develop their assessment tasks they assess both the students’ understanding and their use of typical subject concepts. Some note that you should not be afraid of using unfamiliar words, since this helps the students learn. Instead, teachers comment that you can always explain these to the students in the assessment situation.

All of the interviewed teachers refer to challenges in the assessment of subject content associated with the use of language. Whereas some claim that language should not be part of the assessment in subject matter courses, a majority acknowledge that it is. One teacher mentions that his colleagues believe assessment must be in Swedish since they are held accountable, meaning that they need to be able to see and hear themselves what the students know. They also feel it is their responsibility to prepare them for future education, where Swedish is required.

The distinction between the *what* and *how* of assessment is not clear cut and it sometimes overlaps. When teachers talk about *what* they assess, they often mention *how* assessment is done at the same time, as in some of the examples above. Several teachers acknowledge that assessment in their subject is based on the students’ written skills to a large extent. While some believe this is how it should be, others state they would prefer only oral assessments. In the section below, the focus is on the tests and assignments used (these are mostly written, since written assessment samples were collected from the teachers at the interviews). The *how* of assessment is displayed in the development and use of different modes and types of assessment, as well as test administration involving accommodations applied at the schools, which is described below.

The “*how*” of assessment—the use and importance of different assessment types

In the survey the teachers were asked about several aspects as regards the *how* of assessment, both in relation to their subject and local policy at their schools. The focus of some items in the survey that will be presented in the following was on *assessment types* and teachers’ beliefs about the importance of different assessment formats for grading, such as oral versus written. Some items had to do with the assessment procedure and accommodations used, some of which included the teachers’ and the schools’ language policy as perceived by the teachers. One survey question asked about test development, another about accommodations used when students take the test (e.g., help from an interpreter or the use of translation apps), and yet another question referred to how they used national tests with NAS. The

findings from the survey introduce this section and then there is a look at interview data and assessment samples to illustrate some of the interviewed teachers’ assessment beliefs and practices.

A question that appeared to be a little provocative, causing reactions from several teachers, concerned which types of assessments carries most weight when grading (Q26). Many chose the option *other* and felt urged to share in the open-ended comment box that they use all types available, and that it varies or depends. Several also mentioned the need to use whatever works to try to draw anything they can from the students, depending on the students’ strengths. Teachers often referred to using complementary oral checks after a written test. However, the question was asked since teachers in the previous study in the CLIL context said they preferred written assessments, stating that these felt more reliable and legally safe (Reierstam, 2015). See Figure 3.6 below for the distribution in this study (N = 185).

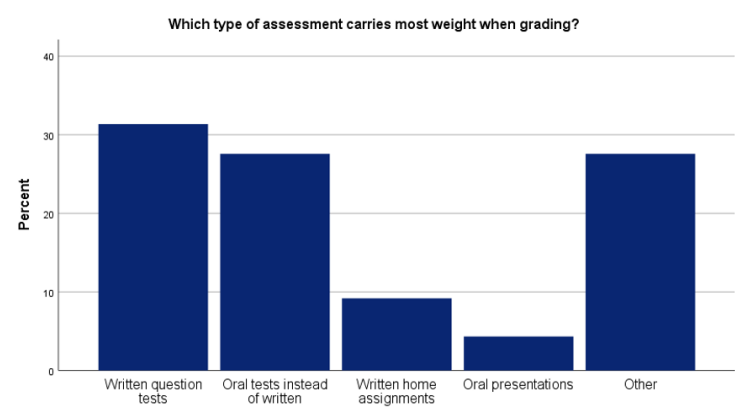


Figure 3.6 Teachers’ attitudes with regard to the importance of different assessment types for grading

The question above referred to assessment formats used for grading among NAS. To enable comparison, the teachers were also asked if the same applied for all students (Q27); the results indicated that the response to the previous question revealed accommodations made for the NAS (see Table 3.20 below).

Table 3.20 Teachers’ perceived importance of different assessment types for grading (%)

	Among NAS (N=185)	In general (N=175)
Written question tests	31	42
Other	28	28
Oral tests instead of written	28	11
Written home assignments	9	14
Oral presentations	4	5

The number of respondents differs slightly in the two groups (NAS vs. in general), which may reflect the fact that certain teachers in the introductory program only work with NAS. The results indicate that teachers adapt their assessment strategies and types depending on what is called for based on the students' strengths and a perceived need for variation; this may help explain the category labelled *other*. Once again, the comment box generated many answers along those lines. The number of teachers using oral tests instead of written is markedly higher among NAS, which can be considered an accommodation. However, written tests are preferred in general, with the number of teachers who favor written tests with NAS also being high.

Looking across subjects and school levels reveals that 39% ($n = 30$) of the NS teachers prefer written tests with NAS compared with 26% ($n = 24$) of the SS teachers; both of these groups include teachers with a dual certification. However, the NS teachers also had a higher frequency for oral tests instead of written: 32% ($n = 25$) compared with 24% ($n = 22$) for SS teachers. The SS teachers instead reported a higher preference for other types of assessment: 32% ($n = 29$) compared with 22% ($n = 17$) for NS teachers. The SS teachers also reported using written home assignments slightly more, although the actual figure is small: 12% ($n = 11$). The same tendency was revealed when looking at teachers' preferences in general, when not necessarily dealing with NAS. Even higher figures, such as 56% ($n = 41$) of the NS teachers reported preferring the use of written tests, and the SS teachers preferred mixing categories and displayed the highest frequency for *other* at 32% ($n = 28$).

Looking at school levels, written tests were preferred by all, but especially in the introductory program, where the figure was 58% ($n = 18$). Here it becomes a bit trickier, since some teachers teach at several levels and are not included in the count. The rest of the distribution resembles that reported for the subjects. When the teachers were explicitly asked if they use the same assessment tasks for NAS as for their other students, 24% ($n = 45$) reported that they do and 41% ($n = 79$) stated that they use partly the same, making accommodations for the NAS depending on language proficiency. Comparing Grades 7–9 in compulsory education with Grades 10–12 in upper secondary shows that the former teacher group makes more accommodations (55%, $n = 58$ vs. 32%, $n = 9$), but the actual count for Grades 10–12 is low. In Grades 7–9, 24% ($n = 25$) state that they use exactly the same assessment tasks, versus 46% ($n = 13$) in Grades 10–12. The introductory program was excluded, since it only works with NAS.

All teachers use and rely on results from written tests to some extent. The question is what kind of tests they use. Do they develop their own, or work together with colleagues? To what extent do they use standardized tests such as national tests and material provided by the NAE? Do they use digital tests or portfolio assessments? Do they use oral and written assessments to the same extent? In answer to question (Q24)—*How often do you use the following when you assess students' content knowledge in your subjects?*—a majority of the teachers reported using self-developed tests, 83% ($n = 159$) of them using these *always* or *often*. Collapsing the *always* and *often* categories, 67% ($n = 127$) use written assessments, 51% ($n = 97$) use oral presentations, and 29% ($n = 55$) use a portfolio frequently. For the question as to whether they use digital assessments, only 18% ($n = 35$) reported doing so *always* or *often*. The assessment types are presented in order of frequency, with the most frequent at the top. (See Table 3.21 below.)

Comparing subject groups showed that 25% ($n = 23$) of the SS teachers use digital forms compared with only 13% ($n = 10$) of the NS teachers. The use of a portfolio was slightly higher among SS teachers (32%, $n = 18$) versus NS teachers (15%, $n = 12$). Adding teachers with a dual-language certificate raises the frequency slightly more for the SS teachers to 36% ($n = 94$) but does not alter the frequency notably for the NS teachers, since there were only 2 teachers in this category (NS plus language). Forty-two percent ($n = 80$) use the current national tests *always* or *often*, with NAS as well as in general. The use

of national tests depends on the subject taught or the school level, since national tests only apply in certain grades and subjects. In upper secondary (Grades 10–12) there are no mandatory national tests in the target disciplines (the natural and social sciences); the NAE provides voluntary assessments in biology, history and social studies. Teachers may also be accustomed to the mandatory tests in other subjects they teach, such as mathematics, Swedish or English, and thus refer to those. The national tests are to be taken in the target language, Swedish. In Grade 9, there are national tests in the natural and social sciences, which are therefore applicable in the introductory program also (cf. p. 55 in Part 1).

Table 3.21 *Frequency in using different assessment types (%)*

	Always	Often	Sometimes	Rarely	Never
Self-developed tests (<i>N</i> = 192)	27	56	17	-	1
Assessment of written assignments (<i>N</i> = 190)	14	53	31	2	-
Assessment of oral assignments/presentations (<i>N</i> = 192)	14	37	41	6	2
The current national tests or assessment material (<i>N</i> = 189)	19	24	37	10	11
Portfolio, collecting students' written texts (<i>N</i> = 192)	14	15	16	22	33
Standardized tests (e.g., old national tests) (<i>N</i> = 193)	7	17	46	21	9
Digital assessments (<i>N</i> = 192)	2	17	23	31	27
Tests developed by others/colleagues (<i>N</i> = 188)	2	13	34	32	19

Collapsing the categories *always* and *often* and comparing NS teachers and SS teachers shows that 53% (*n* = 42) of the NS teachers and 32% (*n* = 17) of the SS teachers use national tests. Adding the teachers with dual-language and subject certificates changes the figures minimally: 35% (*n* = 32) of the teachers with SS and language certification use national tests. In Grades 7–9, 46% (*n* = 105) reported using national tests *always* or *often*, and 43% used them *sometimes*. In the introductory program, 26% (*n* = 9) reported never using them, compared to 5% (*n* = 5) in Grades 7–9. In the introductory program, 40% (*n* = 14) answered *always* or *often*, while 46% (*n* = 48) stated the same in Grades 7–9.

The “*how*” of assessment—local language policy and accommodations

One question in the survey was about policy surrounding national tests: (Q34) *What policy do you have for newly-arrived students and national tests at your school?*

76% of the respondents ($n = 175$) reported to use national tests with their NAS. Of those who take the tests, 45% ($n = 79$) take them without any accommodations, as shown in Table 3.22 below. The most common accommodation used is extended time, as the NAE prescribes. Twenty-three percent ($n = 41$) let the students use an English version, which goes against the recommendations of the NAE. However, the use of other languages is recommended in other contexts. Guidelines shared by the NAE state that the students’ strongest language can and should be used initially, but not with national tests (cf p.55). English is most likely not the students’ strongest language but might be stronger than Swedish in some cases. The local procedures and accommodations used with national tests are presented in the table below in order of frequency, with the most frequent at the top.

Table 3.22 Local procedures with national tests

<i>Newly-arrived students...</i> (<i>N = 175</i>)	Frequency (%)
are given more time	49
do the tests as usual	45
may use a dictionary	41
get words or questions translated by a mother tongue instructor or another bilingual teacher	35
do not take national tests	24
may take an English version of the test	23
take certain parts of the tests	18
may choose if they want to take the tests	18

The question is how the recommendations to use the students’ strongest language translates into practice in assessments other than national tests at the schools under discussion. As regards local policy at the schools, the teachers were asked if they believed there to be any local language policy guidelines for assessment of NAS and whether this is explicitly stated and in that case by whom (Q35). Table 3.23 below presents perceived language policy in assessment at the schools in order of frequency, with the most frequent at the top.

Most of the time there appears to be no explicit policy, according to the teachers (see the last column to the right). Some interesting figures appear in the blue boxes above concerning what is perceived as an implicit policy of using English to bridge the language gap, and what can be considered two opposite approaches of either advocating the use of Swedish or the students’ strongest language if there is a teacher who knows the language. This then becomes a matter of access depending on the students’ L1, “if there is a teacher who knows the language.” In a decentralized Swedish school, it is apparent that policy is either up to the individual teachers or, in some cases, up to the schools. It is also clear that even if a policy applies concerning the use of the students’ strongest language, it is less clear how this should be assessed and by whom (rows 3 and 4). All of these differing approaches can be found to affect the validity and fairness of assessment across schools.

Table 3.23 *Perceived language policy in assessment at the schools (%)*

	Explicit policy between colleagues	Explicit policy from school leadership	Explicit policy from municipal level	Unspoken /implied policy	No policy
Assessment should always be done in Swedish, regardless of the student's proficiency (<i>N</i> = 166)	10	11	1	19	59
Assessment can be done in the student's strongest language if there is a teacher who knows the language (<i>N</i> = 165)	10	9	1	21	59
Assessment can be done in English (<i>N</i> = 167)	8	10	2	21	59
Assessment should always be done in the student's strongest language (<i>N</i> = 168)	7	11	1	13	68
Assessment done in the student's strongest language is assessed by mother tongue instructor /bilingual teacher (<i>N</i> = 166)	4	10	1	7	80
Assessment done in the student's strongest language is translated into Swedish and assessed by me (<i>N</i> = 164)	5	7	1	10	77

In a comparison made between school types, differences appeared: the non-response rate was between 14% and 16% to these questions. The total number of respondents for each school type was: Grades 7–9, *n* = 93–95, introductory program, *n* = 28–30, Grades 10–12, *n* = 24. The teacher groups belonging to more than one school type were not included in this count. The low figures obviously mean that frequencies in percentages sometimes involve very few cases, a factor that must be kept in mind.

Concerning the sub-question *assessment should always be done in the student's strongest language*, 58% (*n* = 55) of the teachers in Grades 7–9 reported that no policy exists, compared with 77% (*n* = 23) in the introductory program and 83% (*n* = 20) in Grades 10–12. Twenty-one percent (*n* = 20) of the teachers in Grades 7–9 claimed to have an implied policy, compared with 4% (*n* = 1) in Grades 10–12 (which obviously represents a very low exact number) and none in the introductory program. This could indicate that the discourse on translanguaging is more widespread among teachers in Grades 7–9 than the rest. Another possible reason might be that teachers in Grades 7–9 more often have a dual teaching combination with the social sciences and Swedish as a second language, or that the organization and collaboration with mother tongue instructors or other colleagues may be more established in the compulsory schools. Twenty percent (*n* = 19) of the teachers in Grades 7–9 stated that there is an explicit policy as regards the use of the students' strongest language either among colleagues or from the school leadership; the corresponding figures were 23% (*n* = 7) in the introductory program and 12% (*n* = 3) in Grades 10–12.

A similar tendency can be seen in the next sub-question, *assessment can be done in the student's strongest language, if there is a teacher who knows the language*. Of the teachers in Grades 7–9, 43% ($n = 40$) reported that no policy exists, compared with 76% ($n = 22$) in the introductory program and 83% ($n = 20$) in Grades 10–12. In Grades 7–9, 33% ($n = 31$) of the teachers claimed to have an implied policy, while only 4% ($n = 1$) in Grades 10–12 and 3% ($n = 1$) in the introductory program did so. In Grades 7–9, 23% ($n = 21$) of the teachers stated that there is an explicit policy, either among colleagues or from the school leadership, while the figure was 21% ($n = 6$) in the introductory program and 13% ($n = 3$) in Grades 10–12.

For the next sub-question concerning the *use of a mother tongue instructor/bilingual teacher in grading and assessing tests done in the student's strongest language*, Grades 7–9 stand out once again compared to the rest. Numbers are high for all three school types as regards *no policy exists*, with 75% ($n = 71$) and 79% ($n = 22$) giving this answer in the first two school types, and 88% ($n = 21$) in Grades 10–12. However, 11% ($n = 10$) of teachers in Grades 7–9 report that there is an implied policy, while the answer was 0% for the introductory program and 4% ($n = 1$) for Grades 10–12. Of teachers in Grades 7–9, 14% ($n = 13$) stated that there is an explicit policy either between colleagues or from the school leadership, while 21% ($n = 6$) answered the same in the introductory program and 8% ($n = 2$) in Grades 10–12. The tendency is the same concerning the reported *use of a translator* so that teachers are able to evaluate assessment results themselves. In Grades 7–9, 16% ($n = 15$) of the teachers report the existence of an explicit policy, while the corresponding number for the introductory program is 11% ($n = 3$) and in Grades 10–12, 8% ($n = 2$). Fourteen percent ($n = 13$) of the Grades 7–9 teachers reported an implied policy, compared with only 4% ($n = 1$) in the introductory program and 0% in Grades 10–12.

When comparing the policy *always use Swedish* in assessments, it is interesting to note that teachers in the introductory program are more prone to state that an explicit policy exists for the use of Swedish: 39% ($n = 11$) report that Swedish should be used, either expressed between colleagues or from the leadership. In Grades 7–9, the frequency is 16% ($n = 15$) and in Grades 10–12, 17% ($n = 4$). Grades 10–12 have the highest frequency for *no policy exists* as regards the use of Swedish, at 75% ($n = 18$), compared with 63% ($n = 60$) for Grades 7–9 and 43% ($n = 12$) in the introductory program. Nineteen percent ($n = 18$) in Grades 7–9 claim there is an implied policy. When it comes to the use of English as a strategy for assessment purposes, Grades 7–9 differ from the other categories: 27% ($n = 25$) of the teachers claim there is an implied policy of using English and 14% ($n = 13$) state that there are guidelines from the school leadership to use English.

To summarize, in cases where some kind of policy applies, Grades 7–9 seems to have more explicit and implied rules than the rest. Since Grades 7–9 represent compulsory school this may not be surprising. Looking at The Ordinance for Compulsory School (SFS 2011:185), existing language policy states that Swedish should be used no less than half of the time and instruction should grant the students an opportunity to acquire an advanced academic language proficiency in Swedish as well as functional English. The only case where the introductory program displays slightly higher frequency for an explicit policy than the rest is concerning the use of mother tongue instructors to assess tests done in the students' L1, which is reported by 21%, although the actual count is small ($n = 6$).

The “how” of assessment—teachers’ language policy and accommodations

While the question above referred to local assessment language policy, another question related to the teachers’ own practice was as follows: (Q29) *How often do the newly-arrived students use the following language in assessment in your subject(s)?*

The respondents were requested to answer all that apply as regards the use of different languages or means of expressions as listed in the table below. Collapsing cases, 90% ($n = 170$) reported using Swedish always or often. Fifty-two percent ($n = 91$) said that the students *never* or *rarely* use their L1 and 72% ($n = 121$) *never* or *rarely* let the students use their strongest language. Twenty-nine percent ($n = 69$) reported using English *always*, *often* or *sometimes*, with the highest frequency for *sometimes*.

Table 3.24 Teachers’ language policy in assessment (%)

	Always	Often	Sometimes	Rarely	Never
Swedish ($N = 188$)	50	40	7	1	2
The student can draw or use other forms of expression (e.g., labs) ($N = 181$)	10	27	25	22	16
The student uses several languages, depending on what works ($N = 178$)	7	14	28	20	31
The students’ mother tongue/L1 ($N = 175$)	1	17	31	22	30
English ($N = 178$)	1	11	27	35	26
The students’ strongest language, other than those previously mentioned ($N = 169$)	3	8	17	20	52

Comparing subjects and collapsing all three categories (*always*, *often* and *sometimes*), the NS teachers use other modes of expression much more frequently than the SS teachers, which probably has to do with the character of the subject: 82% ($n = 81$) of the NS compared with 46% ($n = 85$) of the SS teachers, including teachers with a dual certificate in language. Teachers in the SS disciplines, and particularly when teachers with a language combination are included, reported more positive answers in regard to letting the students use their strongest language. Collapsing cases for *always*, *often* and *sometimes* showed 55% ($n = 46$) for SS teachers and 28% ($n = 21$) for NS teachers. On the other hand, when it comes to letting the students mix languages, 34% ($n = 30$) of the SS teachers responded negatively, compared with 22% ($n = 17$) of the NS teachers.

In the survey, the teachers were asked what kind of adjustments they make when they develop the assessment tasks (Q31); see Table 3.25 below, with responses arranged in order of frequency. The most frequent accommodations provided by the teachers when they develop assessment tasks and tests are visual support (such as pictures, for instance), use of simplified language, and more open assessment types (where the students can express themselves) as opposed to the closed formats offered in selected

response items; the latter are used less frequently by the teachers, at least as an accommodation. The SR format might be of help, since less language production is required of the students; on the other hand, high quality multiple-choice questions are difficult to develop. At the bottom of the list of possible accommodations are accommodations related to the students' strongest language and potential translanguaging practices.

Table 3.25 Accommodations provided by teachers' in the assessment tasks (%)

	Always	Often	Sometimes	Rarely	Never
Pictures or other visual support (<i>N</i> = 184)	17	46	28	6	3
Using simplified language (<i>N</i> = 188)	14	30	34	13	9
Open format where the students can express themselves (<i>N</i> = 185)	6	35	42	15	3
Including a word list in Swedish (<i>N</i> = 187)	9	24	33	19	16
Using model texts/templates (<i>N</i> = 185)	7	27	33	18	15
Using formats that require less language production (e.g., multiple choice) (<i>N</i> = 186)	4	27	40	19	10
Providing start-up words/sentences (<i>N</i> = 183)	5	21	26	24	24
Tasks are translated to the students' L1/strongest language (<i>N</i> = 187)	2	16	26	18	39
Including a word list in another language (<i>N</i> = 183)	4	10	21	18	47

When the teachers in the survey were asked which accommodations they make in the assessment situation (*Q33*), the accommodations most frequently reported to be used *always* or *often* were extended time, 76% (*n* = 143), and use of Google Translate, 57% (*n* = 106). Other language-related accommodations had slightly to much lower frequencies: use of interpreter/mother tongue instructor, 33% (*n* = 62); use of multiple languages as long as the students make themselves understood, 33% (*n* = 62); and use of L1/strongest language, 21% (*n* = 39).

Following this investigation of teacher beliefs as expressed in the survey with regard to preferred assessment types, accommodations, language use and test development, the next section presents teacher beliefs as expressed in the interviews and teacher practices as seen in their self-constructed assessment samples. All of these are intended to discern what the teachers' assessment practices look like with NAS.

The “how” of assessment—examples from the interviews and the assessment samples

In the interviews, teachers express various beliefs as regards what item types work best based on their experiences of different assessment formats and accommodations with NAS. In the following, the teachers’ beliefs and self-reported practices are presented together with examples of test items from their self-developed written exams and assignments used for assessment purposes.

In the development of written question tests, the teachers often acknowledge that item types represent different levels of knowledge. Hence, they state that SR items can only be used to assess very basic conceptual understanding, as shown previously in the section on what teachers believe is assessed.

Below is an example from a test in biology. The series of digits below the question show that it can only generate E scores, as represented by the number “3.” (Numerals in the second and third slots would have indicated the possibility of A or C scores.) This procedure is often used by teachers for scoring.

<p>Example 1: Biology, Grade 9</p> <ul style="list-style-type: none">Match the concepts A–C with the correct definition 1-3: <table><tr><td>A. Top consumer</td><td>1. Zoo plankton</td></tr><tr><td>B. Producer</td><td>2. Sea eagle</td></tr><tr><td>C. Consumer</td><td>3. Zostera marina (seagrass)</td></tr></table> <p>3/0/0</p>	A. Top consumer	1. Zoo plankton	B. Producer	2. Sea eagle	C. Consumer	3. Zostera marina (seagrass)
A. Top consumer	1. Zoo plankton					
B. Producer	2. Sea eagle					
C. Consumer	3. Zostera marina (seagrass)					

Another kind of item format often used is to have students state only one or two words, as in the example below from the same teacher. As can be seen from the numbers underneath, this question can generate E and C scores, but not A scores (E/C/A).

<p>Example 2: Physics, Grade 9</p> <p>1. Mention one or several appropriate forms of energy that may explain these events:</p> <ul style="list-style-type: none">a) A ball rolling on the groundb) Pulling back a sling shotc) A bouncy ball bouncingd) An applee) A hydropower plant producing energy <p>(5/2/0)</p>
--

The teacher uses the same assessment for NAS as for the other students, but the NAS are allowed accommodations such as using translation apps or oral follow-up. The teacher explains that the goal is for every student to be able to write a test in Swedish.


The test item below uses a picture, which can be considered a kind of scaffolding; however, this is necessary for this type of question. This question can generate grades across the whole spectrum, even

the highest grade, A, depending on the quality of the student’s answer. The question word *why* also implies higher-order thinking skills.

Example 3: Biology, Grade 9

- Which of the organisms in the picture most likely contain(s) the most toxins? Explain why!

1/1/1



Although the teachers note that it often is challenging for the NAS to both understand and respond to CR questions, since they require more language production, several claim that it is easier for NAS to produce answers on their own in a simple CR format than in response to SR questions. This correlates with what teachers reported in the survey (cf. Table 3.13). The teacher below mentions both the open question format and other accommodations that work with NAS.

T9 I think open questions work very well, [and] to talk with the students, have oral tests or get help from the Arabic-speaking teacher for students who don’t know that much Swedish, but it is really difficult.

The example below illustrates an open question used by the same teacher in a test in geography. The test is used for all students and the given concepts are thus provided for everyone. The question is at the top, followed by an entire blank page where the students are to write the answer (which can generate grades along the whole scale, E–A). The question is assessed holistically and is one of three test items comprising five questions (two items are made up of a pair of questions). All questions on the exam require lengthy answers. No pictures are used, but one of the questions includes population charts.

Example 4: Geography, Grade 8

- Why do people move to cities or other countries and what are the consequences?

Use these concepts in your text: **emigration, immigration, urbanization, sparsely populated areas, densely populated areas, pull and push factors**

(E–A)

The topic in this test item can be supposed to relate to the NAS’s own experiences, which may be an advantage if they have something they want to share. However, it may also stir uncomfortable emotions

that could contribute to bias in the test compared to their peers. How and if this might affect the assessment depends on how different responses are valued on the scale; i.e., are personal experiences an advantage? No benchmarks or alternative responses were discussed as regards the perceived qualitative levels.

A similar question can be found in another teachers' written exams for the introductory program, as seen in the example below. Unlike Example 4, no words are provided, but it explicitly addresses the student's ability to compare his or her own experiences. This is one of seven test items on the exam. No grading criteria or benchmarks are provided.

Example 5: Social Studies, Language Introduction
<ul style="list-style-type: none"> Do you think there are good laws in Sweden? Why? Is there anything that is better in your country? Discuss.

In relation to the test above the teacher is asked how questions are graded:

- R How do you assess? It doesn't say here if it is a C- or an E-question or how do you weigh it?
- T5 You look at the knowledge requirements, if they have very good knowledge [...] You start from the course plan, "has very good knowledge", well then they have grasped a lot of the core content, and they have been able to develop, that is how. The tricky part is when you develop these tests, to be able to make questions that will generate answers that will tell if they have reached a higher level.

Once again, the knowledge requirements are mentioned and the intended learning objectives as expressed in the syllabi. Teachers often refer to how they balance the accountability requirements of curricula with the special measures they need to take in consideration of the students' needs. Even when they do their best to find reliable assessment tools, some of them note that the NAS do not respond or behave as expected in relation to the teachers' ordinary frame of reference. The NAS respond differently than native Swedish students, which is challenging when the same measure is used with all students. The same is claimed as regards the benchmarks in the national tests: they are not as applicable with NAS. Yet some teachers mention that they use the national test format as a guide, and even copy questions from old national tests but make alternations. Half of the test items in the national test generally use an SR format. Teachers acknowledge that it is difficult for them to construct valid SR items themselves:

- R Do you ever use multiple-choice items?
- T5 It is so hard to make good multiple-choice questions, so no, I don't.

The difficulty in constructing SR questions thus prevents some of the teachers from using them as accommodations for NAS, but the teachers still seem to believe it is best to use CR question formats that serve the NAS well. One solution many teachers advocate is to use open questions, which give the students a lot of freedom to express themselves while at the same time providing the right scaffolding for the question. This can be compared with what is suggested in Cummins' matrix (cf. Analytical models in Part 1). One of the teachers mentions how she balances the accountability perspective in

assessment with a learner perspective by telling the students to write anything they have learned before handing in the exam, even things that have not been asked for on the test. According to her, this allows students who are struggling to feel they can offer something, and the teacher gleans at least some information.

A very common procedure is to use oral follow-up or oral alternatives when the students fail to perform on written tests:

T9 That’s what I find is so good with oral assessment: you notice how much they understand, and you can ask follow-up questions.

In the survey, as reported previously, the results indicate that short written assignments are perceived as easier for the NAS than oral assignments. However, this also depends on what is perceived as an oral assignment. One teacher in language introduction says that he can tell from the way the NAS are reasoning and asking questions during class what they know and if they have read the literature. The survey mentioned “prepared oral assignments,” which require more individual performance from the students compared with oral dialogic assessments, which are more scaffolded by the teacher’s follow-up questions (as noted above).

Still, teachers’ assessment is based on what the students write, to a large extent, even though some express that they would rather use other formats such as only using notes about performance in class or oral assignments. Several comment that the NAS themselves want written tests and sometimes require more challenging tasks, explicitly saying that they “don’t want the easy ones.” One teacher refers to the use of questions from the course book, letting the students send in their answers via a digital platform and then receive comments from the teacher. The questions represent certain grade levels already established by the textbook:

R So this first question represents E-level?
T1 Yes, but if you do a little bit more you can reach C, you can attain a higher level [...] but it is really the questions from the course book that I have used here. [...] I return them with comments, and that is how I assess.

One of the questions is shown in Example 6 below. No specific grading criteria or rubric are provided. The questions are assessed holistically and based on how much the answer has been developed and if the students try to think outside the box to attain a higher grade (as in a C), according to the teacher.

Example 6: Social Studies, Language introduction
<ul style="list-style-type: none">• What is the difference between how individual states are governed and how the world functions? (E–C)

When the students are making calculations in physics, language is not needed in the actual calculation, but the students need to both understand the question and be able to explain their calculation process. One of the physics teachers explains that sometimes it is hard to judge if it is a matter of a student's misinterpretation of the question or a bad calculation. He mentions an example that could generate up to a C that involves a teeter-totter; in the question, it says that two boys sat down "at the tip." One of the NAS had not taken the placement into account when making his calculation and drawing the required figure. If you don't understand the placement on the teeter-totter, it alters the entire calculation. When it comes to the word teeter-totter, the students could have asked what it means, as they could with "at the tip," according to the teacher. He says he also wants some advanced items that make the students think a little bit more to show proof of a more advanced level, as in the item below, which can generate C- and A-scores:

Example 7: Physics, Grade 11

- In order to study the surface of the planet March a satellite is to be placed in an orbit around the planet, so that the orbit equals the planet's rotation around its own axis. Calculate the satellite's distance from the surface of March. Use necessary data from the formula (astronomic data) to perform the calculation.

0/2/3

For the higher grades, the teacher says that it is important that students show that they can draw conclusions and make their own reflections to describe or explain a phenomenon. When asked if this is something they do in class, he confirms "all the time." He also mentions going over all the correct answers after the exam, thereby turning it into a learning opportunity for the students. He gives the students the chance to make clarifications after the test and sometimes even awards them an extra half credit as encouragement if they show good reasoning skills. The assessment is learner-focused, although the knowledge requirements are at the center and are driving the assessment practices.

One of the SS teachers points to a test item on Sweden's governance that seemed to be challenging for all of the students, not just the NAS:

Example 8: Social Studies, Grade 8

- How does the process work from proposal to decision? For example, a ban on cell phones in school.

This test item does not provide any scaffolding, and the students must remember the steps to be able to describe the process. The same teacher explains that they now have a focus on language development in all subjects, and the SS teachers have started to integrate lists of linking words in the exams for all students to use. Examples of these include words to express cause or reason (*since, due to*); words to contrast and make comparisons (*on the other hand, as opposed to*); and words to elaborate (*in other*

words, which implies). In a similar test item on a different test on law and order for Grade 7, he has provided highlighted keywords and given instructions as to how to organize the answer (providing the students, in other words, with certain scaffolding).

Example 9: Social Studies, Grade 7

- When a crime is committed you can say that a process begins. Describe and explain the entire process, starting with a committed crime to when a person is sentenced. (**Arrest, detention, prosecution, trial, verdict**).

Show how **police, courts** and **the penal system** cooperate/are connected.

It is good idea to organize your answer in steps and try to use appropriate concepts.

As an example of a challenging type of item, he mentions a question on democracy from a national social studies test:

T13 This is about democracy and it comes from the Agency, and it is a little tricky for many students; especially those with a different first language, simply because there is a lot of text and you have to figure out what to write. Many students find that difficult, to start to write and understand everything. Here they are supposed to write as much as they can about democracy.

The difficulty consists in the amount of reading that is required of the students as well as a lack of directions on how to structure the response. One of the SS teachers points out that she is not supposed to assess the coherence of the students’ texts, but rather their thoughts; therefore, she must ask herself if she is assessing subject content or language, especially since language is not 100% functional at all times and thus risks getting in the way. Even if she uses the same tests as her colleagues who do not have NAS, she, unlike them, has to make accommodations before, during and after the test. Her students are used to the oral format, and sometimes beg to do it orally when she is pushing them to write:

T3 A lot of students say: “But I can tell you,” and I respond: “Yes, you can, but I want you to write first, you need to practice taking a written test, and then we can do it orally as well, so you can show your knowledge properly.” Many times, they cannot show their knowledge in written form, but I think it is important for them to practice the test format that is used in school.

The teacher, who is a certified SSL teacher, explains that it is a delicate balancing act to give students the accommodations they need while at the same time trying to prepare them for reality, such as taking a test within a given time frame and doing written tests, which she says they will need when they get to upper secondary. When asked about progression and what is required for a higher level, she mentions that if a student shows the ability to explain what happens after a crime, for instance, or during a lawsuit, it is not impossible to reach the A level. One of the questions in her exam on law and order includes the description of a short scenario about a girl, who is the victim of a crime, when she is raped while walking home from a friend’s on a Friday night:

Example 10: Social Studies, Grade 8

- Read the story and answer the questions.

Sara (aged 17) was on her way home from a friend's place Friday night at 10:00 pm. As she was walking through a park, she met a group of boys from her school. They started talking, but she continued walking. One of the boys (Johan) grabbed her arm and forced her to stay.

[...]

Afterward, Johan walked away, leaving Sara in the park. Sara called her mom to tell her what had happened.

- a) What happens next? Answer as extensively as possible and try to use the correct concepts. (Start from where the story ends up until the perpetrator receives his punishment.)
- b) Which punishment do you think is appropriate in this case? Motivate why.
- c) What are the consequences of this crime? (Think of as many perspectives as possible.)

This is an example of a question that could generate grades along the entire scale, from E to A. At this point in Grade 8 the students rarely reach the more advanced grade levels, the teacher says, since they have not had the time to acquire the necessary skills. Sometimes the teacher helps out as their secretary. The example above is one of nine items on the test; several, like this one, include sub-questions. Reading is required, but the NAS can get help from the teacher since lack of understanding or reading ability should not be allowed to get in their way, according to her. The original text of the test item has been shortened in this example. In relation to the topic of the story and the students' background, it is worth reflecting on whether and how this potentially could interfere with the students' ability to perform, even though the topic may be perceived as engaging and authentic. Sensitive content is sometimes mentioned as a threat to fair and valid assessments and may be particularly pertinent in the social sciences, where the purpose sometimes is to elicit argumentative skills with the use of engaging topics.

When it comes to accommodations, issues with finding a good balance have been mentioned. How much and what types of accommodations can be used to make valid and fair assessments? The teachers mention a number of accommodations that are used in the development of the actual tests, during the tests and after: formulating easier questions, translating tests into the student's language, granting permission to ask for clarification during the test, the use of translation apps, offering extra time, carrying out oral follow-up and providing supplementary assignments. In integrated environments in regular classes the same tests are used for NAS as for the rest of the students; the newly-arrived, however, can do complementary oral tasks or follow-up with the teacher to make clarifications, and extra time is granted to all students in need. When asked about accommodations that are considered when the test questions are developed, one teacher responds:

T5 I like to add a picture. You work with the concepts, but I often flip the question, so they don't have to learn the concepts, but they need to explain what they mean. [...] I can relate to my own experience and think, I don't always remember the concepts, but I can explain, and that is what knowledge is about. [...] When I develop a test I tend to think that if they [first] get a word they need in order to explain, and then they come to the explain-questions where they need to use these concepts, and they already have them [the concepts] earlier in the test, they can use that as a support.

The teacher above assumes that it is easier for the students to produce an explanation than remember the words, since this is something they have done in class. Other teachers assume that the students need both the concept and the explanation, and therefore use matching, as seen in Example 1 above. Several tests have at least some items that consist of SR items. One test is made as a rubric where the students can choose an alternative by marking the correct box. The teacher who works in language introduction says that she sees very clear progression in the students' proficiency after Christmas, although not at the necessary pace to finish a course within the expected time frame. However, she notes that she can switch from more scaffolded formats to test items that require more writing and language production. In an exam on law and order during the fall semester she uses question types that require very little language production. One of the seven items includes 13 concepts that the students are requested to use in sentences. As she claimed in the quote above, this is intended to provide them with the concepts as scaffolding (see Example 11 below).

Example 11: Social Studies, Language introduction
<ul style="list-style-type: none">• Use the concept correctly in a sentence. Police: <i>The police arrested the man who had abused the old woman.</i> a) Witness: b) Suspect: [...]

Some teachers state that they would much rather use only oral assessment or classroom assessment if it were up to them. Three factors are expressed as hindering them: first, the students often prefer written tests; second, several teachers mention that they need to prepare the students for higher education, and therefore the students need to practice using written tests; and third, there is not enough time and there is no organization in place that would allow them to sit down with all the students for oral assessment, making written tests easier to orchestrate.

Since language represents an obstacle for students in demonstrating their content knowledge, the question is whether the students must use Swedish or if they can use their strongest language:

T6 [B]oth. I want answers in Swedish, but instead we have translated the questions to their first language, their native language, so they can read the question in their language and Swedish, and then they answer in Swedish; but they could use translation apps.

The teacher in the example above wants the students to write in Swedish but believes the questions may be given in the students' language to help them understand. If their Swedish skills are insufficient, they

can use an app. One of the other teachers has a different strategy, partly because she has colleagues who know the language of the students:

- T3 They can use their own language when they are answering questions on a test. If they don't have the words, they can write in Arabic. Most of them speak Arabic, and that is an advantage since I have a colleague who speaks Arabic and he can help me translate. Those who know English can use English.
- R Is there an explicit policy in your school?
- T3 No, I have done it for a long time, and then colleagues have started to, because I have told them it is not that complicated; many seem to think it is hard.
- R What do you mean, not so complicated? Can you explain?
- T3 Well, there is often someone in the school who speaks Arabic, Dari or Somali, and they are willing to help, if I ask ahead of time. If there isn't anyone, I can always copy and run it through Google Translate just to get an idea of what the students are trying to say, and then they can tell me, but it is to show them that you are interested in knowing what they know, even if it is not a hundred percent right.

The teacher, who is an SSL teacher, points to the pedagogic reason to show interest in the students' background and to find out what they know, even if it is not part of the intended learning outcome. The follow-up question concerns how this gives any input in relation to the knowledge requirements in the objectives:

- R If you use Google Translate or a colleague's translation, how can you say something about their ability to reason and discuss and so forth? How can you see any progression, or is it mainly a matter of a passing grade?
- T3 Well, it is difficult to attain a higher level, C or A, if you don't know enough language; unfortunately, that is the case, but it is before you know the language that you need to use your language, to learn to reason and reach the more advanced levels.

Translanguaging is thus perceived as a temporary accommodation, an intermediary phase on the road to the intended target language, Swedish. The teacher above mentions that on assessments she often gets a lot of mixed languages: some in Swedish, a word of Arabic here and there, some more in Swedish. At this level, Grade 8 she thinks it works, but states they need to know Swedish to attain higher grades and be prepared for upper secondary, where there will be other language requirements. One of the other teachers mentions how she once translated a test to the students' first language, but the student preferred to take a simplified version of the test in Swedish.

- T5 Sometimes I get help from the study counselors, in a test situation for instance. [...] I have wonderful colleagues there, and [the students] can write in their own language and the study counselors will translate it for me, so they can show their subject matter knowledge. Sometimes the study counselors are present when they take a test, [...] but it can be a challenge if they don't know the concepts in their own language. [...] If they have a small basic vocabulary in their own language, the time of schooling in Sweden will become really long.

Even with help from study counselors or translators, the students with poor language skills in their L1 will still be at a disadvantage. The students' backgrounds will have an impact on their progression and the prospects of attaining a grade within the regular time frame. Below, under concluding remarks, the interviewed teachers' suggestions as to how to get the NAS to pass and how to strengthen assessment are presented.

In summary, the teachers believe language is required of the students, particularly in relation to higher-order thinking skills in order to demonstrate more advanced knowledge. The teachers are often aware that they can provide the students with key words in the assessment items or give them guidance on how to structure their answers. Oral solutions can substitute for poor written skills, but only for a time, since written skills will be needed later. The interviewed teachers generally believe assessment should be done in Swedish, at least when the students become proficient enough. Use of translators, apps and mother tongue instructors are seen as temporary adjustments until the students can use Swedish. If the goal is for the NAS to know Swedish and be able to communicate subject content in Swedish, how do the students get to that point? How does this affect how teachers perceive of the language used in class, and how much they focus on teaching different language to the students? One of the research questions asked how teachers' assessment practices are affected by their beliefs about language in education, as in curricula, their subjects' syllabi, translanguaging practices and local organization. According to constructive alignment theory, both instruction and assessment should reflect the intended learning outcomes. In this section, teachers' perceptions of the *what* and *how* of assessment shows that language is an integrated part of assessment, although incoherent language should not be perceived as incoherent content knowledge, as stated by one of the teachers.

The teachers express confidence in their assessment skills for the most part, but acknowledge that it is tricky, particularly with NAS. They also mention that the assessment situation sometimes becomes like a "try out," where you realize afterward that certain questions did not work as planned, or that the NAS do not follow the common pattern. The teachers are very faithful to the syllabi when it comes to what to assess in their subjects, but several acknowledge that they see language development in their subject as an area of professional development.

Comparison across teacher groups

In this section, findings in relation to Research Question 3 are presented with a comparison of teachers' assessment beliefs and assessment practices based on subject identity and school level. Some variation appeared in the cross tabulations, which was presented above. In order to find out if there are statistically significant differences between independent teacher groups, non-parametric analyses of variance were conducted using Mann-Whitney and Kruskal-Wallis tests, where larger values indicate larger differences. The significance level was set at $p < .005$.

Findings from the Mann-Whitney analyses comparing NS teachers and SS teachers are presented first, followed by results from the comparisons between SS teachers with and without a dual teaching identity in a SS subject and a language (i.e., SSL, EFL or Swedish). Thereafter, results from the Kruskal-Wallis test comparing teachers from three school levels (i.e., Grades 7–9, the introductory program and Grades 10–12) are presented.

The focus below is mainly on statistically significant results, but some interesting non-significant findings are noted. For an overview of all results, see Appendices I–K.

Comparing NS and SS teachers

To perform a bivariate analysis using the Mann-Whitney U test, a new variable was computed containing only two collapsed categories: NS teachers and SS teachers. Teachers with a dual combination with a language subject were thus integrated into the major categories. Significant differences were found

across multiple questionnaire items and dependent variables, and therefore some are presented in synthesis with other questions in the tables in order to make the tables more manageable. In Table 3.26 below, significant differences are presented as regards teachers’ reports about the use of oral assignments in class and how they perceive their level of difficulty in assessment.

Table 3.26 Comparing NS and SS teachers’ use of assignments and perceived language difficulty

Variable	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
Scale: Never–often (Q19), Easy–difficult (Q20–21)					
(Q19) How often did you let the students do a prepared oral assignment in a group?	NS	80	77.81	4 295.500	.023
	SS	91	93.20		
(Q20) How difficult do you think the following is for your NAS: ...oral presentation after preparation	NS	82	95.99	4 468.000	.025
	SS	92	79.93		
(Q21) How difficult do you think the following is for your NAS: ...everyday language	NS	83	98.75	3 009.000	.008
	SS	93	79.35		

It is interesting to note the internal correlation in the responses within the subject groups: the SS teachers state that they do more prepared oral group assignments in class, while they also perceive of oral assignments as easier compared with the NS teachers. The question is whether they do oral group assignments because they are perceived as easier, or if these get easier because practice makes perfect—or maybe both. In the interviews, teachers often perceived oral assignments as accommodations to bridge language barriers and facilitate dialogic negotiation of meaning. Table 3.26 also shows significant results concerning everyday language, which NS teachers find more difficult for the NAS than the SS teachers. As discussed previously, this judgment also depends on what is considered everyday language. Some of the NS teachers mentioned in the interviews that it is challenging for NAS to access experiments, for instance, due to a lack of experience of everyday activities such as bicycling or teeter-totters. In this case, everyday concepts become crucial in order to access particular phenomena. Another way of interpreting everyday language may relate to filler language used as a substitute for more academic language. Either way, as noted previously, some of the SS teachers acknowledged the same challenges for NAS in the interviews as the NS teachers in terms of everyday language due to poor exposure outside of school.

In Table 3.27 below, statistically significant differences in relation to perceived difficulty for NAS in the use of various subject skills and function words often used in test items are presented. In all cases, the NS teachers report a higher perceived level of difficulty than the SS teachers. It is interesting to note that the verbal functions *name* and *state* are perceived as more difficult by the NS teachers than the SS teachers. These two terms are typically used in selected response item formats where less language production is required, representing lower-order thinking skills.

Table 3.27 Comparing NS and SS teachers' perception of difficulty in subject skills

Variable Scale: Never–Always	Categories	N	Mean rank	Mann- Whitney U	Asymp. Sig. (2- sided)
(Q22) How difficult do you think the following is for your NAS: ...describe	NS	83	97.06	3 066.000	.016
	SS	92	79.83		
(Q22) How difficult do you think the following is for your NAS: ...discuss	NS	83	95.92	3 160.500	.035
	SS	92	80.85		
(Q22) How difficult do you think the following is for your NAS: ...explain	NS	83	99.05	2 984.000	.005
	SS	93	79.09		
(Q22) How difficult do you think the following is for your NAS: ...name	NS	82	100.22	2 811.000	.001
	SS	93	77.23		
(Q22) How difficult do you think the following is for your NAS: ...state	NS	83	96.99	2 989.000	.011
	SS	91	78.85		

The NS teachers report higher in relation to the importance of students' ability to use a subject-specific genre, as seen in Table 3.28 below, whereas the SS teachers report significantly higher on the importance of subject-specific skills and the use of cross-disciplinary academic language. This may possibly reflect the character of the subject disciplines, with the use of laboratory reports, for instance, required in the NS, while the SS may require more generic academic literacy skills.

Table 3.28 Comparing NS and SS teachers' perception of the importance of language features

Variable Scale: Not important–very important	Categories	N	Mean rank	Mann- Whitney U	Asymp. Sig. (2- sided)
(Q28) How important is the following: ...the students can use subject skills, e.g., to argue, draw conclusions	NS	81	90.59	4 407.000	.022
	SS	94	94.38		
(Q28) How important is the following: ...the students can use a subject-specific genre	NS	80	106.85	2 132.000	.000
	SS	93	69.92		
(Q28) How important is the following: ...the students can use cross-disciplinary academic language	NS	81	80.17	4 441.000	.032
	SS	94	94.74		

Table 3.29 below summarizes statistically significant differences in the use of different assessment formats.

Table 3.29 Comparing NS and SS teachers' use of different assessment formats and types

Variable <i>Scale: Never–Always</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
(Q24) How often do you use the following with your NAS: ...standardized tests/old national tests	NS	82	101.76	2 766.500	.001
	SS	94	76.93		
(Q24) How often do you use the following with your NAS: ...national tests	NS	81	97.81	2 850.500	.006
	SS	92	77.48		
(Q24) How often do you use the following with your NAS: ...digital tests	NS	82	78.99	4 552.000	.022
	SS	93	95.95		
(Q24) How often do you use the following with your NAS: ...oral assignments	NS	81	78.32	4591.000	.013
	SS	94	96.34		
(Q24) How often do you use the following with your NAS: ...written assignments	NS	80	77.03	4 517.500	.007
	SS	93	95.58		
(Q24) How often do you use the following with your NAS: ...a portfolio	NS	81	79.63	4 485.000	.036
	SS	94	95.21		

As already noted in the section containing the descriptive statistics, the NS teachers seem to rely more on national tests, see Table 3.29 above, whereas the SS teachers report using more digital tests. The SS teachers use more varied assessment formats, such as oral and written assignments, where the latter represents written essays and reports as opposed to regular tests containing questions. This may also explain the difference in the use of portfolios, which are used more by SS teachers compared with NS teachers. However, the NS teachers reported higher usage than the SS teachers for other types of assessments; while this result was not statistically significant, it does show that almost half of the responding NS teachers reported using other types of assessments (see Appendix I), which may consist of laboratory assignments, for instance.

Table 3.30 below relates to accommodations in the administration of tests and how teachers prepare students before an assessment.

Table 3.30 Comparing NS and SS teachers' use of accommodations

Variable <i>Scale: Never–Always</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
(Q29) How often do the students use the following in assessment: ...Draw or use other forms of expression	NS	81	99.44	2 151.500	.000
	SS	85	68.31		
(Q31) Accommodations used when you develop your assessment items: ... Wordlists/explanations of words	NS	80	77.19	4 425.000	.018
	SS	92	94.60		
(Q31) Accommodations used when you develop your assessment items: ...Sentence starters	NS	76	69.26	4 730.000	.000
	SS	93	97.86		
(Q31) Accommodations used when you develop your assessment items: ...Model texts	NS	78	70.22	4 780.000	.000
	SS	92	98.46		
(Q31) Accommodations used when you develop your assessment items: ...Open questions where the students can express themselves	NS	78	72.91	4 648.000	.001
	SS	93	96.98		
(Q32) Prepare the students for an assessment: ...Teach typical subject vocabulary and concepts	NS	80	80.91	4 287.000	.031
	SS	94	93.11		
(Q32) Prepare the students for an assessment: ...Write texts together in class	NS	80	74.90	4 608.000	.003
	SS	92	96.59		
(Q32) Prepare the students for an assessment: ...Present oral models	NS	80	73.41	4 807.500	.001
	SS	93	98.69		

The Mann-Whitney test showed a significant difference in the use of other forms of expressions and letting the students draw, for instance (see Table 3.30), which reportedly is more frequently used in the NS. A possible explanation for this, as before, is in the character of the subjects, since it is more common to draw figures and models in the NS disciplines. In all the other items, including differences in the use of accommodations in the development of test items, the SS teachers reported significantly higher results. This may be explained in part by some of the teachers' dual certification in a language subject, as investigated below. The examples expressed in the survey items are typically found in the teaching and learning cycle (TLC) (Gibbons, 2002, 2018), which was mentioned at the beginning of Part 3, and in Part 1 to be used as scaffolding with language learners. In the same vein, the SS teachers display significantly higher results in relation to their preparation of the students and their teaching of all three language features (Q32) listed at the bottom of Table 3.30.

Comparing SS and SS + language teachers

Since the group of teachers with a dual combination with language was particularly big in the social sciences, yet another variable was computed for a second bivariate analysis containing only two categories—SS teachers versus SS and a language subject (e.g., Swedish, SSL or EFL)—to find out if the dual certification and language teaching experience has an effect on teachers’ reported practices. Below, the focus is on statistically significant differences; complete results are presented in Appendix J.

Table 3.31 below shows statistically significant differences between the two groups: SS versus social sciences and a language subject, henceforth called SSLang.

Table 3.31 Comparing SS and SSLang teachers’ use of assignments, perceived difficulty in assignments and language genre

Variable	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
Scale: Never–Often (Q19), Easy–Difficult (Q20–21)					
(Q19) How often do you let the students do a prepared oral assignment individually?	SS	58	43.14	1 471.000	.006
	SSLang	39	57.72		
(Q20) How difficult do you think the following is for your NAS: Oral presentation after preparation	SS	58	53.82	851.500	.031
	SSLang	39	41.83		
(Q21) How difficult do you think the following is for your NAS: Everyday language	SS	58	54.38	877.000	.031
	SSLang	40	42.42		

Above, it was noted that the SS teachers use more prepared oral assignments than the NS teachers (cf. Table 3.26); here in Table 3.31, it is apparent that it is specifically the teachers with a dual combination including a language subject who do this. For a related question, the results are significantly different as regards the SS teachers’ perception of oral presentations being more difficult. Hence, the question is: do the SS teachers without a dual-language combination use prepared oral assignments to a lesser degree because these are perceived as difficult, or are they perceived as difficult because the SS teachers have less practice?

The results for the perceived difficulty of everyday language follow the same patterns as above, with the SS reporting that they perceived it as more difficult than their SSLang colleagues. In the section above comparing NS and SS teachers, it was the NS teachers who reported significantly higher results compared to the SS teachers in relation to the perceived difficulty of oral assignments and everyday language.

The rest of the findings in relation to the use of various assessment formats and accommodations follow the same pattern, with the SSLang teachers displaying similar differences when compared with their non-language colleagues as with the NS teachers compared with the SS teachers above. As regards the use of standardized tests or old national tests (Q24), the SSLang teachers show significantly higher

results than the SS teachers (Table 3.32 below). In the comparison above between NS and SS teachers, it was the NS teachers who reported significantly higher results in relation to the same question, whereas for the frequent use of a portfolio, the SS teachers displayed more significant results compared with the NS teachers.

Table 3.32 Comparing SS and SSLang teachers' use of assignments, accommodations and perceived importance of different language genres

Variable <i>Never–Always (Q24), Very important–Not important (Q28)</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
<i>(Q24) How often do you use the following with your NAS: ...standardized tests/old national tests</i>	SS	59	43.07	1 589.000	.002
	SSLang	40	60.23		
<i>(Q24) How often do you use the following with your NAS: ...a portfolio</i>	SS	59	45.09	1 469.500	.034
	SSLang	40	57.24		
<i>(Q28) How important is it that the students are able to use cross-disciplinary academic language?</i>	SS	59	45.73	1432.000	.043
	SSLang	40	56.30		
<i>(Q31) Accommodations used when you develop your assessment items? ...Wordlists or explanations</i>	SS	57	43.75	1 382.500	.036
	SSLang	39	55.45		
<i>(Q31) Accommodations used when you develop your assessment items? ...Model texts</i>	SS	58	42.11	1 472.500	.004
	SSLang	38	58.25		

In the bivariate analyses between SS and SSLang teachers, the latter group shows significantly higher results than their SS-only colleagues. This could mean that the more “language teacher,” the more use of different writing assignments and old national tests or standardized tests. A similar significant difference is found as regards the perceived importance of cross-disciplinary academic language, where the SSLang report higher than their SS colleagues but, as seen above, the SS teachers report higher than the NS teachers.

It is interesting to note that among all of the nine suggested accommodations that can be used in test development, only two display significant differences between the two groups of teachers (i.e., *including wordlists or explanations* and *model texts*); see bottom of Table 3.32 above. In both cases, the SSLang group reports significantly higher results compared with their non-language colleagues.

The final table, Table 3.33, presents differences between SS and SSLang teachers in how teachers prepare the students for an assessment. The SSLang teachers display significantly higher results in relation to all language activities in class—activities that coincide with suggestions in the TLC model

(Gibbons, 2002, 2018), as expressed above. The only activity that does not display significant differences relates to subject-specific vocabulary and concepts (cf. Appendix J, Question 32). Again, the language teacher identity seems to have an effect on the reported beliefs and practices.

Table 3.33 Comparing SS and SSLang teachers' preparation of students for assessment

Variable	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
(Q32) How do you prepare the students for an assessment? ...Look at useful sentence structures	SS	58	42.64	1 500.000	.004
	SSLang	39	58.46		
(Q32) How do you prepare the students for an assessment? ...Study the structure of typical texts	SS	58	39.24	1 755.000	.000
	SSLang	40	64.38		
(Q32) How do you prepare the students for an assessment? ...Provide written examples/text models	SS	59	40.27	1 754.000	.000
	SSLang	40	64.35		
(Q32) How do you prepare the students for an assessment? ...Write texts together in class	SS	59	39.90	1 658.000	.000
	SSLang	38	63.13		
(Q32) How do you prepare the students for an assessment? ...Present oral models	SS	59	42.15	1 584.000	.001
	SSLang	39	60.62		

Below, results are presented from the comparison of teachers across different school levels.

Comparing teachers across school levels

For the descriptive analyses, seven groups of teachers were identified due to the fact that some teachers work across multiple school levels. However, for this analysis, three individual categories were used in order to facilitate the analyses of variance and have groups containing bigger samples. Hence, the teachers with intersecting school levels were not included ($N = 167$). When the Kruskal-Wallis statistic was calculated to determine whether there was any statistically significant difference between teachers at different school levels, significant differences were found in relation to a number of variables. In order to find out where the significant difference lies in the Kruskal-Wallis test between the three groups, the Bonferroni pairwise post-hoc test was employed ($p < .005$). A selection of significant results that appear to be of special interest for the overall comparison in the study are reported thematically in the following. All results, significant as well as non-significant, are presented in Appendix K.

Activities and preparation in class

The first theme presented in relation to school levels covers frequency of different activities in class and activities used as preparation for assessment. As above, when analyses of variance were conducted in relation to teachers' subject identity, the Kruskal-Wallis test used for the multivariate analyses of school types revealed significant differences as regards the use of prepared individual oral assignments in class ($\chi^2 = 13.76, p = .001$). There was strong evidence of differences between Grades 7–9 and 10–12 and between Grades 7–9 and the introductory program. Hence, this reveals that the use of prepared individual oral assignments (*Q19*) is more common in the introductory program and in Grades 10–12 at upper secondary. A significant difference was also found in terms of letting the students complete short individual assignments during class, which happened least frequently in Grades 7–9 ($\chi^2 = 6.44, p = .040$). As for all the other activities mentioned in relation to oral and written assignments used in class individually or in a group, no significant differences were found related to school level.

In terms of preparation in class before assessment (*Q32*), there were significant differences in relation to the study of typical texts together in class ($\chi^2 = 13.76, p = .001$) when adjusted for using the Bonferroni correction. The significant difference was found between Grades 7–9 and the introductory program. The Kruskal-Wallis test showed that the study of texts is least common in Grades 7–9 (cf. Appendix K). None of the other activities displayed any significant differences due to school level.

Perceived difficulty and importance of different language activities, registers and skills

A second theme covers perceived level of difficulty for NAS in different activities and language registers. Comparing teachers' reported beliefs showed significant differences in relation to three of four oral and written activities (*Q20*). The only activity that did not generate significant differences was longer written assignments, which was perceived as difficult across all school levels. There were significant differences regarding the category of oral discussions in class between Grades 7–9 and 10–12, as well as between Grades 7–9 and the introductory program ($\chi^2 = 18.65, p = .000$); the same was true for oral presentation after preparation ($\chi^2 = 13.23, p = .001$). Both of these activities were perceived as most difficult in Grades 7–9. This correlates with the results in the previous section as regards frequency, since oral assignments are perceived as more difficult and reported less frequent in Grades 7–9. Short written assignments were also perceived as most difficult in Grades 7–9 and less difficult by each school level ($\chi^2 = 22.65, p = .000$). The challenges for teachers in Grades 7–9 become apparent, since all activities involving both oral and written skills appear to be more difficult at this level. On the other hand, for NAS who have reached upper secondary, the perceived difficulty seems to decrease and the question then is whether the teachers' experiences in Grade 10 are based on NAS who are "new newcomers," or if they come from instruction in Grade 9 and are therefore better equipped.

Significant differences were also evident in relation to perceived difficulty in most of the language skills exemplified in *Q21*. For the question about how difficult subject-specific vocabulary and concepts are for the NAS, significant differences were found between Grades 7–9 and the introductory program ($\chi^2 = 8.59, p = .014$). Teachers in the introductory program reported subject-specific concepts to be the least difficult of the three groups. In terms of subject-specific skills, significant differences were found between Grades 7–9 and Grades 10–12 ($\chi^2 = 7.77, p = .021$). As regards subject-specific oral genre, significant differences were found between Grades 7–9 and Grades 10–12 ($\chi^2 = 10.13, p = .006$); this was also the case for subject-specific written genre ($\chi^2 = 12.45, p = .002$). All three of the language genres reported above were perceived as most difficult in Grades 7–9 and then less difficult by each

school level. Comparing results for everyday language, significant differences were displayed between Grades 7–9 and the introductory program, where it was perceived as being the least difficult ($\chi^2 = 15.22$, $p = .000$).

Looking at perceived difficulty in all the subject-specific skills and verbal functions exemplified in Q22, the perceived level of difficulty was highest in Grades 7–9 across all function words; however, not all differences were statistically significant using the Bonferroni post-hoc correction. The verbal function *argue* was perceived as less difficult by each school level and significant differences were found between Grades 7–9 and Grades 10–12 ($\chi^2 = 19.50$, $p = .000$); this was also true for *analyze* ($\chi^2 = 7.82$, $p = .020$). Significant results were found for *reason* between both Grades 7–9 and the introductory program, and between Grades 7–9 and 10–12 ($\chi^2 = 10.68$, $p = .005$); it was perceived as most difficult in Grades 7–9. The ability to *make comparisons* was perceived as the most difficult in Grades 7–9 and significant results were found ($\chi^2 = 8.85$, $p = .012$). To *show connections* was equally perceived as less and less difficult by each school level, and a significant difference was found between Grades 7–9 and 10–12 ($\chi^2 = 7.72$, $p = .021$).

Comparing the perceived importance of different language registers showed significant differences only in relation to one of four school-related language features: cross-disciplinary academic language. The results revealed that teachers in the introductory program perceived this as particularly important compared to the other school levels, and significant results were found when comparing Grades 7–9 and the introductory program ($\chi^2 = 8.00$, $p = .018$). The teachers in the introductory program reported significantly higher use of portfolios for assessment ($\chi^2 = 12.68$, $p = .002$) (Q24), which may correlate with the perceived need for general academic language.

Language and accommodations used for assessment

Comparing teachers' reported use of different languages for assessment (Q29) revealed significant differences across school levels in the use of Swedish ($\chi^2 = 31.76$, $p = .000$) and the use of the students' L1/mother tongue ($\chi^2 = 14.96$, $p = .001$). Use of Swedish was significantly more frequent in Grades 10–12, whereas the reverse was found in Grades 7–9, with significantly more frequent use of the students' L1. The same tendency was found in the translation of assignments to the NAS's L1 or strongest language (Q31), with a significantly higher frequency in Grades 7–9 ($\chi^2 = 26.72$, $p = .000$). This also correlated with results regarding accommodations used in the assessment situation (Q33), where a significant difference was found for Grades 7–9 and more frequent use of the students' L1 when responding ($\chi^2 = 25.96$, $p = .000$) or the use of a mix of languages, as in translanguaging ($\chi^2 = 13.20$, $p = .001$). Other significant results were found in relation to language accommodations, with a significantly higher frequency for Grades 7–9 compared with the rest.

In contrast, it is interesting to note significant differences in relation to teachers' reported practices in giving corrective feedback to the NAS regarding language (Q30). Three of four alternatives in relation to this question suggested that corrective feedback is given on language in written assignments and in assessments. Teachers in Grades 10–12 reported significantly higher on all of those compared to the other school levels (see Appendix K). Grades 7–9 reported significantly lower in relation to all, and as for providing corrective feedback on language in class, the introductory program reported significantly higher results ($\chi^2 = 19.24$, $p = .000$).

To sum up this section comparing teachers across three school levels, the results show that whether based on students' progression or not teachers perceive that the challenges across different language

registers are most prominent in Grades 7–9; in a similar vein, the use of accommodations such as the students' L1 or strongest language is most common at the same school level. In relation to the verbal skills that appear at all achievement levels in the syllabi, to *reason* for example, attainment becomes difficult for NAS at most grade levels, particularly the more advanced above E-level. Giving corrective feedback on language is more common at the higher levels, which may not be surprising considering the reportedly more frequent use of other language resources in Grades 7–9. However, for assessment purposes, questions remain as to what is and can be assessed, especially in Grades 7–9 due to the integrative nature of subject content and language. Consequential validity is another concern, depending on whether there is alignment between classroom practices and assessment administration. In the introduction of Part 3 it was noted that 72% of NAS were not eligible to apply to upper secondary school in 2018 (National Agency for Education, 2019e), and this represents entrance into vocational programs where the requirements are slightly lower than for programs with a focus on university studies. The findings presented above indicate that more work remains so that more students can access and engage with high-level subject matter content. In order to become proficient users of the new language, NAS need to be exposed to language features and practice both identifying them and using them in increasingly complex tasks. The NAS who became eligible for upper secondary (28%) have apparently succeeded, and numbers have gone up (compared with 16% in 2016) but the challenge is how to support and help equip the rest of the students.

Below, teachers' self-reported needs for professional development in relation to assessment of NAS are outlined.

Professional development in assessment and language teaching methods

In the survey, the teachers were asked if they thought they needed professional development in assessment of NAS (Q43); 44% ($n = 84$) said *yes*, 17% ($n = 33$) responded *partly*, and 38% ($n = 73$) said *no*. In Table 3.34 are a few examples of what teachers stated in the open-ended comment box. Not all the comments revolve around the teachers' own professional development, but probably reflect what they feel they need in order to ensure valid and fair assessments. When the teachers were asked if language development in their subjects was included during teacher education (Q38), 65% ($n = 125$) responded *practically none*, 18% ($n = 34$) said *a little*, and only 7% ($n = 14$) responded *quite a lot*. The reports were quite evenly distributed across teachers of all subjects. On the other hand, when asked if they have received any professional development during their time as teachers (Q39), 37% ($n = 71$) responded *yes*, *pretty often*, and 40% ($n = 77$), *on some occasions*. This indicates that professional development in the field has increased lately, which correlates with the increase in material and courses on the topic which was mentioned in the introduction.

Table 3.34 Teachers’ self-reported professional development needs

Q43 Examples of professional development needs:
<ul style="list-style-type: none">• More explicit policy guidelines and clear directives for assessment• Coordinate/adapt the social sciences and SSL• Language development across the curriculum• How to assess content knowledge rather than Swedish• What to assess? The way the knowledge requirements are formulated language is required• How to structure the questions and perform reliable and valid oral assessment with the mother tongue instructor even if I don’t know the language• Benchmarks showing texts at E level• How to handle assessment in heterogeneous groups with only a couple of NAS• Assessment of language progression in longer written texts

At the interviews, the teachers were asked when and where they had learned about assessment; in addition, they were asked whether they had learned about language development, and if so, how. Most stated that they learned about assessment after and not during teacher education. The teachers mentioned that they had learned from their own experience, from co-assessment, from experienced colleagues, and through online groups and courses. As regards language development, some mentioned SSL colleagues or collegial training, and some mentioned their own SSL teacher education.

One teacher commented on the difficulty in performing valid assessments as an inexperienced teacher and the importance of co-assessment for reliability reasons, referring to when teachers sit together and discuss the assessment of national tests:

T10 [A]nd then you see the discrepancy between different schools and teachers [...] in religion we had someone who put an A, and others an E, so...

R Was that an open question format?

T10 Yes it was, makes you think, what happened there?

R Could you agree on something? [...]

T10 We had a discussion [...] and we had to agree, so I guess we landed on a C or B or something. But it has to do with experience, one of us, the colleague who said E I think, was a recent graduate, and she was the outlier in the group. I believe recognizing E and A has to do with experience too; after grading a few tests it is different.

Being in an environment with language learners means that language inevitably becomes more visible, raising the question: what does this do to the teachers’ perceptions of themselves as language teachers or language providers? The answer clearly varies, and may depend on how they perceive of the concept of a “language teacher”:

- R Do you see yourself as a language teacher?
- T6 No, not at all! No, in that case I rather see myself as a natural science teacher with a little... [it's] more like I encourage them outside of class. I see myself as an extra mentor, which I think is very important, telling them to speak Swedish, practice Swedish; and [I] tell them they are doing great, because their working morale is rarely a problem.

Another teacher is asked if he feels that it is rewarding to focus more on language and work with the students' language development in class:

- T13 Yes, it is very rewarding I must say, simply because you develop your subject in a different way, it becomes very didactic, how to teach. [...] I understand everything [in my subject] but it is easier to work with language development to help the students understand too, to read and write, or whatever it is.
- R Would you say that you become more of a language teacher?
- T13 Yes, a little maybe, well, yes, if you work with language development you do what a language teacher does in a sense, absolutely, but I think all teachers should do it in that case, one might think that everyone needs to be some kind of language teacher.

A teacher with dual certification in Swedish as a second language and social sciences refers to how language teaching methods merge with her teaching strategies in the subject content course:

- T12 I used my SSL skills to advance more in social studies. I read aloud to the students before we started working, for instance; I never left them to themselves to do that.

Two of the social science teachers in the introductory program talk very enthusiastically about their teaching methods with the newly-arrived. One of them is self-critical when asked how he can go from teaching concepts to illustrating how to use them in discussion as well:

- T4 That is not my strength, I confess, it is not. My instruction is very relational, [...] but the counselors gave me credit for being good at explaining and filling in the gaps, so I must have done something right. But I think it stays with introducing the concepts, drawing and illustrating while writing the word next to the picture, talking about connections, cause and effect; well, there I did work with higher-order thinking skills.

He notes that if he were going to continue teaching at the introductory program next year, he would have taken a course in language development methods that the municipality offers in collaboration with a university. Unfortunately, the introductory program is shrinking, so he is going back to regular classes. Another teacher explicitly states how working with NAS means that she has developed as a subject teacher:

- T5 I am so much more responsive now to the language developing methods, I do a lot more explaining of concepts than I did previously, and it is more obvious to me to focus on practicing skills [...] I have also got new perspectives compared to when I had more homogenous groups with native Swedish students.

One of the teachers who himself has a different L1 than Swedish points out that for reasons of equity and giving the NAS equal access to future opportunities, he must help them with the language and correct their mistakes, even though he is a subject matter teacher:

T8 Language is very important, and there are at least two Sweden when it comes to language, and the sad part is that no one will correct their language mistakes which characterize someone as lower class if the teacher doesn't do this at an early stage, so I think I should.

The teacher in the last example is the only teacher who explicitly adopts a class perspective and claims that it is his responsibility to equip students to provide them with equal opportunities.

Below are some concluding remarks in relation to the findings; subsequently we will delve into the discussion in the final section of Study II.

Summative comments about teachers' assessment among NAS

What are teachers' assessment beliefs and practices regarding NAS, and how have these been shaped? When teachers were asked in the survey if they find assessment difficult in the NAS context (*Q42*), 34% ($n = 65$) responded *yes*, 47% ($n = 88$) *sometimes*, and 19% ($n = 36$) *no*. A majority (81%) thus acknowledges that it can be challenging. As for the 19% who said *no*, it would have been interesting to hear more about what they have to say. However, questions arise as regards the challenges, and if and how these can be ameliorated. The special needs of the group of language learners are present, but what about the conditions provided by the schools and teachers' training and professional development? In other words, what can be done?

The comment field in the survey generated many complementary responses covering different aspects; these are summarized below and organized according to emerging themes (see Table 3.35); an expressed need for more research and literature as regards assessment in multilingual assessment and issues with the organization in the schools, for example lack of collaboration with colleagues. The students' background is described as a challenge for various reasons. Although the word validity is not explicitly used, the teachers point to threats associated with the appropriateness and validity of inferences and interpretations that can be made from test scores. It is described in terms of access to resources that are not normally permitted (e.g., the teacher providing answers), or referred to as language that would be considered to represent construct-irrelevant variance.

The teachers show great awareness of the difficulties involved in multilingual assessment and the comments exemplify their collective experiences. Even if many of the teachers in this study have a fairly long background with second language learners (cf Figure 3.4), they mention a need for training in test development; how to construct alternative tasks that will support the NAS to show their understanding and analytical ability. The teachers also express a need to develop their own interpretative skills in order to better discern students' understanding in spite of limited communicative skills and if possible be able to separate students' content mastery from their current language proficiency. See Table 3.35 below for a summary.

Table 3.35 Teachers' comments on the major challenges of assessing NAS's content knowledge

Identified themes:	Teachers' comments:
Lack of research and input	<ul style="list-style-type: none"> • "Too little has been said and written [on this matter]."
Issues with the organization	<ul style="list-style-type: none"> • Lack of time • Lack of study counselors • Lack of translators • Lack of collaboration with colleagues
Issues of validity in interpreting results	<ul style="list-style-type: none"> • "To interpret their knowledge in spite of insufficient language." • "Difficult to know if it is the language or the content knowledge [that is lacking]." • "Hard to tell what they know [by themselves] if I give the answer and they nod in agreement. There is always uncertainty." • "Uncertainty [as to] what the student did on his/her own. Can they use their cell phone, computer, mother tongue instructor?"
Student background and heterogeneity of the groups	<ul style="list-style-type: none"> • "The groups are not based on educational level but on what schedule works for the students." • "The students want to give an impression they have understood, but they haven't." • "Newly-arrived students often have a lot of pressure from home which makes them underperform out of nervousness. Moreover, the educational traditions from their home country make them focus on factual recall and they don't do well on discussion topics."
Issues with language/content and format of assessment	<ul style="list-style-type: none"> • "To construct assignments where the language isn't a hindrance to showing content knowledge." • "[Being able] to assess analytical skills, the student's ability to draw conclusions and make connections in history."

Issues in relation to language in assessment were by far the most frequent subjects of comment. The lack of language skills is preventing understanding, but more so production, as in the last point in the table above concerning the difficulty in assessing analytical skills when a student's language skills are limited. A reflection from one of the participants in the survey points to many issues related to the choice of language for assessment and the intended learning goal (translated from Swedish):

It is difficult to know how much emphasis should be on disciplinary language and concepts in Swedish. In our school, we believe that the student should have the right to show content knowledge in the language he/she masters, but when the student comes to upper secondary, they must show their knowledge in Swedish. This means that a student who has been assigned a good grade in a subject won't pass when he/she comes to upper secondary. So, the question is if it really is an advantage to use the strongest language in assessment when it doesn't go well in the next step of education. (NAS subject matter teacher, Grades 7-9)

Consequently, there is an ongoing balancing act between what is best for the student and concerns about accountability. Teachers have dual and sometimes contradictory professional demands that are pedagogically driven in relation to the students, and accountability driven in relation to the teachers' professional authority.

Discussion

The purpose of Study II was to contribute to research on assessment in a multilingual context among newly-arrived students, NAS. In the introduction in Part 1, it was stated that equity and fairness are key features of quality education and cannot be considered in isolation from curriculum, assessment and educational opportunities for the students. Based on the understanding that validity and fairness in assessment are closely intertwined, the discussion that follows will raise questions in relation to both. Kane (2010) posits that validity and fairness represent closely related ways of looking at the same question, i.e., that the intended uses of test scores are appropriate for all over a range of contexts, since “equity in opportunity to learn material is also a validity issue if the purpose of assessment is to evaluate how much has been learned in a particular educational program” (2010, p. 179), that is, in relation to predefined target learning outcomes. An assessment thus risks being inappropriate and unfair to those who have not had an opportunity to learn the skills or material covered within it (Gee, 2003).

In the beginning of Part 3, it was noted that according to several OECD surveys (OECD, 2015b, 2017), grading practices differ between schools in Sweden; the National Agency for Education (NAE) (2015) also notes varying practices and conditions across schools as regards the education of NAS. Stobart infers that fair assessment is inseparable from fairness in access opportunities and in what the curriculum offers (2005, p. 278). The aim here, therefore, was to investigate teachers’ assessment beliefs and assessment practices in a bi- or multilingual context, especially in relation to the role of language when preparing students and planning for assessments in which content knowledge is to be demonstrated in a second language. The preconception was not that the teachers’ self-reported assessment practices would be inappropriate or unfair, but that the teachers’ beliefs and experiences of a complex, multi-dimensional issue—subject matter assessment in a second language—are important to understand in order to make empirically-based decisions in the future.

Assessment in a multicultural and multilingual setting is a balancing act, weighing students’ different backgrounds and equal right to education while aiming at providing equal and comparable grades, although “equal” is not necessarily the same as equitable or fair. Kane (2010) makes a distinction between *procedural fairness* and *substantive fairness*. Procedural fairness requires all test takers to be treated in essentially the same way, taking the same or equivalent tests under the same conditions, or equivalent conditions, and requires that their performances are evaluated using the same rules and procedures (e.g., national tests). Substantive fairness requires that score interpretation and use of assessment are appropriate across groups, which may require different adjustments for groups or individuals. As noted in Part 1 on accommodations and scaffolding, this may cause certain controversies in relation to the appropriateness of different types of accommodations. Do these accommodations offer legitimate support, or do they imply unwarranted support for the groups in question, contributing to skewed results and thereby making them incomparable—a threat to both content and construct validity.

Against this backdrop, with reference to the challenges of the assessment conditions, I will here discuss and draw some conclusions in relation to each of the three research questions outlined in the beginning of Part 3. Thereafter, some concluding remarks are offered, including implications for assessment and policy. Limitations of the study are presented at the end, while suggestions for future research are addressed at the end of the synthesis in Part 4.

Assessment beliefs and the purposes of assessment

- *What are the subject content teachers' self-reported assessment beliefs and assessment practices regarding newly-arrived students?*

Beliefs have long been considered to permeate the perceptions and guide the practices of teachers, and therefore investigation of the teachers' self-reported beliefs in the current context was perceived to be a valuable source of information about the teachers' assessment practices. Moreover, assessment is claimed to hinge upon the knowledge and value systems of educators (Banks, 2004; Scott et al., 2014; Tierney et al., 2011). The teachers were expected to offer honest reflections, although there is always a possibility that participants chose answers based on how they wanted to be perceived. However, as noted previously (Part 1, p. 28f), possible inconsistencies between professed beliefs and observed practices must also be kept in mind (Pajares, 1992). Beliefs as an indicator of behaviors may be filtered through contextual factors as individuals interact with the surrounding world (Li, 2013; Poulson et al., 2001; Tamimy, 2015). At the start of this investigation, there was an expectation of finding both similarities and differences between contexts and subject disciplines depending on the teachers' experiences and background, and therefore the teachers' varying backgrounds were noted, particularly in relation to the third research question.

The conceptual framework leans on key assumptions found in validity theory, assessment literacy and language integration models (cf. Part 1). Assessment literacy is sometimes used to describe teachers' professional knowledge and skills, and a model was used for this study that includes a series of questions—*why? what? how? and and?*—where the last refers to the consequences and inferences of assessment, which is closely related to the broader concept of validity (Cronbach, 1971; Kane, 2006; Messick, 1989). The assessment literacy model has provided a framework for the survey and the interview guide and is therefore used in the discussion of teachers' assessment beliefs below as well, together with questions regarding validity and fairness as outlined above.

As regards contextual factors, a majority of the teachers reported having fairly extensive experience with second language learners: only 14 % had no previous experience. The teachers can thus be said to have a certain background in the field of language learners, in many cases represented by mixed groups where they needed to make special accommodations for NAS as opposed to other students. The teachers reported having on average 3–5 different languages in their classes and a majority (excluding the teachers in language introduction who have only NAS) stated that they have a few students in each class. Teachers were asked if they thought assessing NAS was difficult; 34% said *yes* and 47% stated *sometimes*, which means that 81% of the teachers acknowledged at least occasional challenges.

Starting with the *why* of assessment, the accountability purpose was very pronounced among the participants in the study. Teachers expressed concerns about reliability and legal justice and worried about the risk of perhaps helping the students too much when providing model texts or examples that NAS could potentially copy. Some even mentioned being “suspicious” when assessing students' written texts and were afraid the learning had become too instrumental. Teachers were also apprehensive about any cues they offered in the oral follow-up used to complement written assignments that had proved insufficient to generate passing grades. They worried that their input might risk invalidating the students' performances: “what is the student's actual level?” In the same vein, some teachers reported that they did not want students to do assignments at home or with mother tongue instructors since the performance would be out of their control, thus implying a threat to validity. The same could sometimes be perceived

in relation to accommodations such as cell phones and translation apps, although teachers found these necessary. The concerns were not as clear among the teachers in language introduction, where a great deal of focus was on helping students obtain a passing grade. The accountability function (ensuring valid assessments) seems to be very rooted among teachers, although a mix is often found, implying a learner focus as well (see Figure 3.7 below). The figure intends to show that most teachers represent a mix of learner centeredness to the left and syllabi centeredness to the right in their self-reported assessment practices. It is important to point out that the teachers' positions along the continuum—here represented by stars—are based solely on my interpretations as a researcher. I perceived that a couple of the teachers were more consistent in their descriptions of the accountability purpose, whereas a couple of others were specifically explicit as regards the purpose of helping the students succeed.



Figure 3.7 The interviewed teachers' assessment profiles

Assessment profiles

The purpose of Figure 3.7 is to illustrate the tension teachers experience, maybe more so in this particular context where they must consider both the special needs of the students (toward the left) and provide valid assessment interpretations in relation to predefined knowledge requirements (toward the right). However, a couple of the teachers expressed a greater drive for accountability and were more prone to seeing the students as responsible for their own learning; the opposite, however, was true for a couple of others. Most teachers displayed examples of both perspectives. The illustrative interview excerpts below provide more examples of the tension between the two, typically shown in the mixed example from T6 below.

Pedagogy driven:

T12 That is why I give them study questions before a test and then go over the responses with them, instead of saying “you have a test.” “Yes, but on what?” “Pages 22 to 32, you guess what will show up.” That is not what knowledge is about, knowledge is about skills, knowledge that we acquire, abilities we can express [...] My role as a teacher is not to hinder my students [and] [...] prevent them from succeeding.

Mix:

T6 In our exams, there is nothing that we haven’t talked about already, everything is exactly what we have taught, that is always the case.

T6 The problem with oral assignments is that you have to be very careful [...] because I write down what they say [...] and I don’t want to put my words in their mouth, but it is not easy when you need to explain the question to them. It is tricky to have a ..., legal [assessment], or legal is the wrong word, but I think you understand what I mean

[about written projects and why they do those in school:]
T6 It is about a need to control so they don’t ask mom and dad

Accountability driven:

- T13 We are trying to implement that here [show sample student responses], [...] it is hard with tests, they are not supposed to see the questions before hand, but we have used them for hand in assignments.
- T7 There are [students] who really study, and we are trying to help of course [...] Those who have a backlog in everything, for various reasons, language might be one, [...] they need something other than 70 retakes. There is such a focus on every student's right to get a new chance to show their knowledge, but what are we doing by giving them all these re-exams, what is done to make them learn?

Assessment and beliefs about students, language and local organization

- *How are the teachers' self-reported assessment practices associated with their beliefs about the students and language policy in education (i.e., curriculum, syllabi, translanguaging practices and local organization)?*

Below, findings in relation to the second research question regarding the teachers' perception of the students, language and organization are outlined in summary, including a discussion about how this is perceived to impact teachers' assessment practices and beliefs.

Teachers identified several challenges for NAS in performance and the attainment of higher grades; most of the time, they had to accept that the focus needed to be on the attainment of no more than a passing grade, according to the teachers. As noted in the introduction of Part 3, NAS represent a very heterogenous group (Bergendorff, 2014; Bunar, 2010; National Agency for Education, 2011) and age and time of arrival have been reported to correlate with school attainment (Cederberg, 2012; National Agency for Education, 2005, 2019e). The teachers in this study note that NAS in their contexts (Grades 7-12) have a lot of catching up to do compared with their peers in relation to academic literacy.

The teachers in the present study seem to hold contradictory beliefs about language, whether it is part of the construct and the assessed domain or not. Language and content are inherently interdependent, as noted by Nikula et al. (2016) and thus integrated in the learning objectives in curricula, (cf. Figure 1.3 in Part 1), but this does not necessarily mean that teachers view themselves as language teachers (cf. the teachers' language integration profiles in Part 4). When it comes to how language is handled and by whom, the teachers report varying beliefs as regards who is responsible. Based on the interview and survey data, the reported variation in what schools are able to provide in terms of study counselors, translators and collaboration between colleagues implies a potential threat to validity in assessment. As noted in Part 1, validity is reduced if results benefit a particular group (Gipps & Stobart, 2009) and fair assessment equals fair access opportunities (Stobart, 2005), as in providing necessary language skills to language learners which is discussed below.

Language-related challenges for assessment

The biggest challenge, as noted by the teachers, is closely connected to subject-specific requirements: subject-specific concepts, subject-specific skills and subject-specific genre were identified as the most challenging for NAS. In relation to the verbal skills, expressed in the function words in the knowledge requirements, the teachers stated that it was particularly difficult for the students to reason, analyze and draw conclusions. This is closely related to the discrepancy teachers expressed between what the students think they need to know, i.e. facts, and what is required in the intended learning outcomes, as in higher-order thinking skills. The teachers related that the students' previous academic background generally implied a focus on factual recall, a cultural difference that the teachers almost perceived as impossible to communicate. All teachers noted issues in relation to language and the students' different perception of knowledge, which contribute to difficulty in interpreting the students' responses, along with a difficulty in perceiving if it is language or content that is lacking.

Whereas translanguaging and the use of the students' combined linguistic resources is advocated for NAS (National Agency for Education, 2012a, 2018g) most teachers referred to a monolingual norm, stating that the sooner the students use Swedish the better. Translanguaging was identified as a possible accommodation that works best in language introduction according to the teachers, and they maintained that NAS need to know Swedish in order to succeed in upper secondary school. The consequential validity of translanguaging in assessment needs to be further investigated: How and if it can be used in assessment and its possible social and societal impact (see discussion in Part 4).

Lack of experience and a frame of reference was mentioned as a disadvantage that not only affected NAS but all students in terms of what they talk about at home, thus pointing to the importance of the compensatory mission expressed in the Education Act (A. Andersson et al., 2015; SFS 2010:800). This also points to equity issues in assessment due to different environmental experiences or unequal access (Gipps, 1994).

Challenges related to the assessment format

Teacher beliefs in relation to the *what* and *how* of assessment revealed challenges associated with different written and oral assessment formats. Longer written assignments and oral discussions in class, which require more language production and often lack scaffolding, were perceived as particularly demanding for the NAS. At the same time, teachers tend to use written assessment to a large extent. Some stated that they would rather use oral assessment but expressed a need to equip students for higher education, where written skills are needed. Oral assessment is used as an accommodation, which raises questions in relation to teachers' skills in assessing oral renderings of subject content and how that differs from written presentations. In language testing, oral assessment is a field of its own, bringing up issues in terms of the validity and reliability of the dialogic co-construction between test-taker and test-administrator—often the teacher—(van Lier, 1989; Young, 1995) or where clarifications can be made and meaning negotiated in peer interaction (Brooks, 2009; Ducasse & Brown, 2009; Fulcher, 2016; Moere, 2012, 2016). As already noted, this was something the teachers seemed to be aware of themselves and were concerned about, once again in relation to accountability. It is also a matter of construct validity, i.e. if the oral skills accurately measure the intended skills, and extrapolation (cf. p.45); if the oral test-results can be extrapolated to a targeted construct. The choice of format relates to substantive fairness; whether the student has been appropriately prepared and equipped with oral skills, and procedural fairness, if there is comparability between written and oral results across student groups.

Crooks et al. (1996) mention threats to the valid use of assessments associated with the administration of tests, depending on whether the students have been given the appropriate coaching before the test. If students are given an oral assignment in an assessment situation, the necessary skills must be practiced and modelled in instruction. The same applies to written assignments; otherwise, the assessment format may imply what Messick (1989) referred to as construct-irrelevant variance. The delivery mode, since it is not part of instruction, is thus to be considered “irrelevant” to what is assessed; however, some mode must be used, and thus also specifically taught. Bias in assessment may refer to biased test items or inequality in access to learning, which in turn may be due to real differences in the opportunity to acquire talent or in the opportunity to demonstrate talent, according to Gipps (1994, p. 125).

In relation to alignment, it can be noted that if subject concepts were perceived as the most difficult language feature for NAS, this correlated with what teachers reported focusing on the most in the classroom. However, more language is needed if the students are to perform in both written and oral reasoning. This is an issue that needs to be addressed, i.e. alignment between the intended outcome, assessment format and instruction, to answer the consequential question *and?* in relation to the interpretation and use of test scores.

Variation in teacher beliefs and practices

- *Do teachers’ assessment beliefs and practices vary depending on school level and subject discipline. If so, how do they differ and on what grounds?*

Teachers’ knowledge of the purpose of assessment, how to use various assessment strategies, assessment interpretation and relevant action-taking define their assessment literacy (Abell & Siegel, 2011; Malone, 2013). In a context with second-language learners, the appropriateness of the assessment strategies and how to appropriately prepare students become part of assessment literacy practices (Popham, 2009b). Special assessment issues are associated with students who are not native speakers of the language of instruction as seen above. Popham (2009a) notes that it is important for all teachers to become conversant with the procedures most suitable for these subgroups of students. However, research also suggests that teachers’ beliefs are associated with background features such as teacher experience, context and subject discipline (e.g. L. B. Barnes et al., 2001; N. Barnes et al., 2015; Harris & Brown, 2009; Lavigne & Dalal, 2015; Shulman, 1986). The aim of the third research question was to investigate the possible effects the teachers’ disciplinary background and teaching context may have had on their reported assessment beliefs, which is especially crucial in a context where teachers’ beliefs need to be harmonized with the special assessment issues regarding NAS.

By conducting statistical analyses comparing different teacher groups, the findings revealed that there were significant differences between teachers from different subject disciplines and school levels. The findings in Study II show that depending on subject and school level, teachers sometimes hold contrasting beliefs about what skills are especially important for the students to master, what is particularly difficult for the students, and what activities and skills need to be emphasized in class. Certain differences very clearly relate to beliefs about language as subject content, as medium of instruction and as a national norm. Other differences relate to what kind of assessment format is perceived to be more appropriate for the subjects’ intended learning outcomes and what

accommodations can be used. However, sometimes the reason for the differing self-reported practices is not entirely transparent, as in the case of the statistically significant difference found in the SS teachers' self-reported more frequent use of digital assessment, a possible area of further investigation for the future. It is important to note, however, that there are many similarities in terms of both teachers' perceptions of language learners as well as reported assessment practices. Below, both differences and similarities in the findings across subjects and school levels are discussed.

Comparing teachers across subject disciplines

Teachers are informed and guided by the national curriculum and its subject-specific intended learning goals; this is stated explicitly in the interviews and the assessment samples, but also demonstrated implicitly in the statistical analyses of the survey data. The results from the statistical tests show that teachers from the SS disciplines report using oral activities to a greater extent than the NS teachers, and the latter group perceive oral assignments as more difficult for the NAS compared with the SS teachers. Even though subjects may have differing disciplinary practices based on what is part of the intended learning outcomes, oral versus written forms of assessment relate instead to different modes of delivery. As one of the teachers claimed in the interviews, "it doesn't say anywhere that it has to be in writing," referring to the students' communication of content knowledge. Perceiving oral assessment as an equal or at least comparable form of communication means that teachers' differing practices may imply a threat to equitable assessment and consequential validity when students' results from oral assessments are weighed against results from written assessments, or results from "poorly administered" oral assessments are weighed against results from "well administered" oral assessments. In other words, when oral assessment is recommended as an accommodation, teachers' oral assessment literacy needs to be considered in order to ensure fair and valid assessments.

As regards assessment format, questions have been raised as to whether selected-response could be an option to alleviate stress related to language production. However, teachers in this study acknowledge that valid multiple-choice items are difficult to create. It is interesting to note that in relation to what Dalton-Puffer (2016) calls CDFs (mostly referred to in this study as language functions and subject skills), the NS teachers find to *name* and to *state* more difficult than the SS teachers. These two function words represent lower-order thinking skills and are usually found in items where less language production is required; however, these items require that students know the words and they offer little context (cf. Cummins' matrix, p. 36f). This can be addressed in alternative assessment formats, however, as noted by some of the teachers in the interviews who stated that it helps NAS if the first questions involve matching with explicit word lists, as this allows them to go back to these questions and use the vocabulary later in the test. A portfolio is an assessment format that is less context-bound but gives the students the opportunity to express themselves more freely. This format is significantly more common among SS teachers. The use of subject-specific genre, as in laboratory reports, is significantly more common among the NS teachers, as are national tests and standardized tests.

Following is a summary of what appear to be significant differences between what teachers from the different disciplines report emphasizing in class, what they perceive of as challenging for their students, what they think is important for the students to know, and what kind of assessment format they prefer.

NS teachers find prepared individual oral presentations and everyday language especially challenging for NAS: it is difficult for NAS to describe, explain, name and state content knowledge. To master a subject-specific genre is perceived as particularly important in their subjects, and students can draw or

use other forms of expression. The NS teachers report significantly higher use of national tests (both the current and old available tests), but almost 50% also chose the option *other*, which may indicate that they have other forms of assessment besides those covered in the questionnaire, such as laboratory assignments.

SS teachers who do not have a dual combination with a language subject report that prepared oral assignments are more challenging than their SSLang colleagues, however SS teachers overall claim to give more oral assignments than NS teachers. SS teachers also mention giving both more oral and written assignments (e.g., essays) for assessment than the NS teachers, and they give significantly more portfolio assessments. The SS teachers score significantly higher on the perceived importance of subject skills and the use of general academic language. SS teachers without a language combination express that everyday language is more challenging for the NAS. The SS teachers have a significantly higher frequency of use across most language accommodations in the development of assignments used for assessment (i.e., use of wordlists, sentence starters, model texts and use of open question formats). In relation to the other options (i.e., assignments translated to the students' L1, wordlists in another language, use of simplified language, use of MC items, inclusion of pictures or visual support), no significant differences were found.

Based on the significant differences from this study, it may be safe to suggest that the SS teachers generally are more “language oriented” in their reported assessment formats, which require more general academic literacy. However, the large group of teachers with a dual-language subject combination must be taken into account. The NS teachers, on the other hand, report needs for more specialized subject-specific knowledge expressed as subject-specific concepts and genre. Subject-specific genre and general academic genre both need to be taught and modelled, which is discussed in the concluding remarks. As regards non-significant results, it is interesting that no significant differences are found between the teacher groups in relation to the use of Swedish, the students' L1 or translanguaging. Similarly, no significant differences appear in relation to giving corrective feedback on language in class or in written assignments depending on subject.

Comparing teachers across school levels

What stands out the most in the results of the statistical tests across school levels is the obvious progression found between Grades 7–9 and Grades 10–12, which may seem logical but should be considered in relation to time of arrival of the students. Most language registers and functions are perceived as more difficult in Grades 7–9 as compared with the rest with the exception of longer written assignments, which are perceived as difficult by all. No significant differences are found in the perceived difficulty of general academic language, which is perceived as fairly difficult or difficult by all. For Grades 7–9, significant results are shown in relation to several of the verbal functions, where all are perceived as significantly more difficult with the exception of some for which no significant differences are reported; these include *to formulate questions*, *to identify*, *to name*, *to evaluate* and *to draw conclusions*, some of which represent lower-order skills.

Concerning assessment formats, only one significant difference is found depending on school level: this relates to the more frequent use of a portfolio in the introductory program, where cross-disciplinary language is also reported to be more important compared with the other school levels. Why this is so and how portfolio assessment is conducted is a potential area of future research.

One interesting and noteworthy difference in the comparison across school levels relates to language policy: significant differences are found both in relation to language use in assessment and providing corrective feedback on language. In sum, it can be stated that whereas Swedish is perceived as the norm in Grades 10–12 and also in the introductory program, alternative language practices are used in Grades 7–9, including use of the student's L1 or strongest language, translanguaging or an interpreter. Corrective feedback on language is significantly more common in both the introductory program and in Grades 10–12 at the upper secondary level. For more details, see Appendix K.

Some of the differences expressed above naturally are related to the proficiency level of the students and the perceived guidelines provided by authorities such as the NAE or professional development courses for teachers, both of which state that it is advisable to use the students' strongest language during the initial phase as well as varied forms of assessment. Consequently, it seems teachers in Grades 7–9 have internalized this advice in their reported beliefs. Moreover, teachers seem to use a variety of assessment formats regardless of school level, since no significant differences appear, and several state in the open comment field that this is the case. From an equity and fairness perspective, the question is what happens in assessment if NAS in Grades 7–9 can use alternative languages, and what does a given grade say in terms of language skills? Is there a gap that must be bridged between school levels, even though teachers in the interviews expressed an awareness that they needed to prepare students by practicing writing in Swedish in order to ensure success in upper secondary school?

Irrespective of subject background and school level, it is safe to say that fair assessment is closely intertwined with language practices and policy, which calls for more explicit unified cross-disciplinary standards and procedures. Unified does not necessarily mean doing the same thing at all times, but refers to providing a shared framework. Fairness in assessment means multiple opportunities to meet standards in different ways (Gipps, 1994). Acknowledging that language is part of the assessed domain and the construct implies identifying the strengths and potential weaknesses of different assessment approaches. Gipps (1994) claims that performance assessments—in this study exemplified in portfolio assessments and laboratory exercises and reports—offer an opportunity to assess higher-order skills and both construct and consequential aspects of validity are high. Gipps (1994, p. 85 f) mentions that face validity (i.e., transparency in what is being assessed) is often particularly good in performance assessment. However, she notes that the cognitive requirements of the assessment tasks must be clearly stated—and, I might add, language requirements—in order to enable teachers to teach to the required skills. From a fairness perspective, reliability in performance assessments has been especially vulnerable, since part of the construct implies a move away from highly standardized procedures (Gipps, 1994; McNamara, 2000). In order for performance assessments to provide valid and generalizable results, certain requirements must be met, such as clear assessment specifications, shared understandings of performance standards, training of raters [teachers] and the possibility of co-assessment (Baker et al., 1993). Bachman and Palmer (2010, p. 219) maintain that without defining the construct in terms of both language knowledge and topical knowledge, it is not possible to interpret test takers' performance and to know whether poor results depend on weaknesses in language ability or in topical knowledge.

Concluding remarks - implications for policy and pedagogy

In the interviews, the participating teachers more or less explicitly came up with suggestions they thought would lead to better education practices for NAS; whether these could lead to valid, fair and equitable assessment is a subject for further investigation. The suggestions can be summarized as three alternative crossroads relating to: (a) teacher education/professional development, (b) organization at the schools and (c) curricula.

Possible adjustments in teacher education refers to equipping all subject matter teachers to be language teachers, which has been suggested as good practice by many, especially in relation to students with a different L1 (e.g. Bjerregaard & Kindenberg, 2015; Gibbons, 2006; Hajer & Meestringa, 2014; Kindenberg & Wiksten, 2017; National Agency for Education, 2012a, 2012b, 2016, 2018e, 2018g, 2012c; Schleppegrell, 2011). In relation to international assessment in PISA, almost 80 teacher educators and professors in Sweden signed an article (Dagens Arena, 2019) advocating mandatory courses in second language learning in teacher education, which has not been the case thus far. Many professional development courses are provided for teachers of language learners, often with the advice to use a variety of measures and methods of assessment. If more (or even all) subject matter teachers were given at least language teacher skills, if not dual certification, this might alleviate some of the pressure for collaboration with study counselors and SSL teachers. As the findings reveal in this study, teachers prefer to collaborate with other subject matter colleagues, arguably because they are more familiar with the subject and course requirements. This argument could be falsified if good practice could be found, combining teachers' different skills. One issue right now relates to statistics concerning the high number of uncertified SSL teachers (National Agency for Education, 2019d); this brings us to the next suggestion, organization at the schools.

Many teachers seemed to agree that there could be more collaboration between colleagues, and that they are "not there yet." The way education is organized today does not promote any extra activities involving teaming up with colleagues for exchange or consultation, according to the teachers. Some teachers who have a dual certification with insights into the way the subject syllabi are structured conclude that there is not much room for it, either. The SSL courses have a lot of course content that needs to be covered, especially since there are also national tests in SSL that the students need to prepare for.

Adjustment at the curricular level would imply either alternating the intended learning outcomes and providing different tracks (sometimes referred to as streaming) or creating an intentionally "slower-paced" program. Individualized study plans are already advocated for NAS during the initial phase, but the same curricular goals are used (National Agency for Education, 2016). This would thus imply taking it one step further, making adjustments in the "measuring stick" (i.e., the subjects' knowledge requirements) as well, which would have implications for the procedural fairness of assessment. All students would not be taking the same tests and their performances would not be evaluated using the same rules or procedures; thus, the students would be treated differently. However, it could be argued that this is being done already, where accommodations make assessment appropriate for all using the same standards, which relates to substantive fairness (cf. Kane, 2010). In conclusion, matters involving both syllabi learning goals and working conditions at the schools would need to be resolved in order to promote collaboration. (These teacher-reported suggestions are presented more in the synthesis, Part 4.)

Assessment literacy has been defined as knowing how to assess, how to interpret the results, and how to apply the results in order to improve learning and instruction (Webb, 2002). The goal of providing

valid fair and equitable assessment brings alignment between intended learning outcomes, classroom activities and assessment strategies to the fore (see Table 3.36 below).

Table 3.36 Alignment between intended learning outcomes, teaching and learning activities in class and assessment

Intended Learning Outcomes (ILOs)	Learning and teaching activities in class	Assessment		
		What	How	Grade
Knowledge of subject-specific concepts , <i>e.g., urbanization, verdict, algal bloom, friction</i>	Focus on vocabulary	Understanding of concepts Use of concepts	Selected response Constructed response	E
Knowledge of subject-specific written skills : <i>e.g., draw conclusions, reason, analyze</i>	Focus on how to use verbal skills in writing , modelling language, e.g., “A leads to B which leads to C”	Use of verbal skills and higher-order thinking skills	Constructed response Written format	E–A
Knowledge of subject-specific oral skills : <i>e.g., draw conclusions, reason, analyze</i>	Focus on how to use verbal skills in speech , modelling language, e.g., “A leads to B which leads to C”	Use of verbal skills and higher-order thinking skills	Constructed response Oral format	E–A
Academic Swedish skills (General academic language)	Focus on content-specific and content-compatible language (e.g., on the one hand...on the other, consequently...)	Higher order thinking skills	Longer texts Oral discussions, presentations	E–A
Basic Swedish to communicate subject content	Use of understandable Swedish and possibly translanguaging ? No correction of language.	“Language not assessed” Subject content	Anything to find proof of subject matter knowledge (<i>What</i> knowledge?)	E–?
Basic L1 to communicate subject content	Making oneself understood, use of all linguistic resources, translanguaging	“Language not assessed” Subject content	Anything to find proof of subject matter knowledge (<i>What</i> knowledge?)	E–?
Academic L1 skills	(Mother tongue instruction?)	-	-	-

(Words in bold in order to highlight alignment between ILOs, classroom activities and assessment format)

Table 3.36 provides an illustration of how practical advice that is currently suggested in material from both the National Agency for Education and other sources for teachers working with language learners and NAS could be applied to offer alignment between the intended, implemented and attained curriculum. The table outlines some of the most commonly stated learning objectives according to the teachers. Conceptual knowledge is mentioned in subject syllabi as something that students should be able to develop and is often mentioned in the introductory text on the aim of the subject, as here in biology (National Agency for Education, 2018c):

Teaching in the subject of biology should aim at helping students develop knowledge of biological concepts, theories, models and working methods.

When it comes to the knowledge requirements the focus is on various verbal skills, and the following descriptions are used for the attainment of different grades, where E is the lowest passing grade and A the highest:

- E**
Simple reasoning
- C**
Developed reasoning
- A**
Well-developed/elaborate reasoning

Looking at Table 3.36, it becomes rather clear that in order to attain the highest grades, the students need to be proficient in academic Swedish (blue-shaded areas). The letters to the right illustrate the levels that could be attained, i.e. the whole spectrum (E-A) or only a passing grade (E), although there is a question mark (E-?) to signal uncertainty whether a higher grade would be possible. However, looking at the knowledge requirements in syllabi it is also questionable if knowledge of concepts is enough for a passing grade, i.e. “simple reasoning”. There is also a question in parenthesis in relation to *what* knowledge is learned and assessed when different languages or registers are used. As mentioned in Part 1 (cf. p.19), how translanguaging can be used in assessment with monolingual learning goals still needs to be settled. Academic language in the students’ L1 is a possible concern for mother tongue instruction and therefore beyond the scope of this study.

Students’ proficiency in the language of schooling is a matter of validity and fairness in assessment. If language is not part of the assessment, as claimed by several subject content teachers, then anyone could attain an A regardless of language proficiency level. It would not matter which language was used, as long as procedural fairness could be guaranteed; everyone is taking the same test and is treated in equivalent ways (Kane, 2010). If language skills are not part of the intended learning outcome in subject content courses, language should be avoided or at least ignored, lest it be a matter of construct-irrelevant variance. If, on the other hand, language is indeed part of the construct and the assessed domain, as the syllabi suggest, exemplified by the descriptors above, then language needs to be both assessed and taught; otherwise, there is a risk of construct under-representation. Grades are used for admission to higher education and, as noted by Kane (2010), the use of tests to make admission decisions is inappropriate and unfair to those students who have not been given the opportunity to learn the required skills.

In summary, the teachers in the study are working very earnestly and ambitiously to help the NAS, but the situation is not easy given the varying backgrounds of the students and the sometimes conflicting messages: the admonition to use the students’ total linguistic resources/strongest language (i.e.,

translanguaging), for example, versus the instruction to help the students become proficient in academic language (i.e., Swedish is the norm). Another example is the admonition to support the students (i.e., help them by modeling language), versus the understanding that teachers are to be held accountable (i.e., make sure the students do not cheat) while at the same time using the same standards and grading policy for all students. As stated in a national report from the British government (2008, p. 57f): “If you want to measure change, don’t change the measure. But the nation does—and should change/update the National Curriculum [...] Even where the curriculum has not changed much, the way it is taught may have changed considerably.” In terms of the Swedish context, the background of the student body has changed—one-fourth now has a foreign background—while the curriculum has not changed, but has been revised, most recently in 2019. It is still language dense and requires communicative skills. Meanwhile, the way the curriculum is taught probably has both changed and not; how it is implemented varies depending on school context, which may be one of the most important issues if the goal is equitable and fair assessment for all. Many teachers were not trained to teach language because they were trained for supposedly homogenous monolingual contexts. Even if the sudden massive influx of migrants slowed down after 2017, the heterogenous and linguistically diverse classrooms remain. During the process many teachers have gained valuable experiences that society should use. As one of the teachers stated in Study II, all students, regardless of mother tongue, profit from more focus on language in education.

Limitations of the study

The current study is limited in both scope and time frame. Results may not be completely generalizable, since the sample was restricted due to the limited ability to gain access to a bigger scope of participants. Moreover, the data collection procedure was constrained by dependency on principals as gatekeepers when disseminating the questionnaire. I had little control over how the participants were asked or selected, although I know that it differed since some principals mentioned sending the link to all relevant teachers, whereas others explicitly said they would only target a few; all of these factors may constitute a threat to the construct validity of the study. Sample bias may apply in the qualitative interviews as well, since teachers taking the survey were self-selected, which means that the sample most likely includes teachers with an interest in the field of assessment. Another imitation in relation to the data collection methods concerns the reliance on teacher reports. Classroom observations would have provided a complementary perspective to the teacher reports but was not feasible within the time frame of the present study.

In retrospect, I probably would have formulated the survey questions differently. Time constraints and lack of prior experience in both developing questionnaires and analyzing quantitative data probably had an effect on the survey. The questionnaire was too extensive, and therefore a shorter version most likely would have contributed to a larger sample; instead, I found myself with a high dropout rate (see the section on methods). Extensive questionnaires may also lead to less reliable results, since respondents become less prone to answering sincerely.

Another challenge in the design of the study has to do with combining the two studies, one of which was a licentiate thesis from 2015. The research questions, data collection methods and analytical tools differ somewhat, but the first study informed the second in a sense and had to be taken into account along the way. Although this caused some constraints, I mostly perceived of the interrelationship as an asset, both in providing important insights in related fields of research and in providing opportunities for cross-fertilization since results from the first study had to be revisited. I have learned so much in so

many areas, yet there is more to explore, learn and delve into (see suggestions for future research on p.232.)

Due to limited research on the topics of both Studies I and II, I was not able to find many relevant references for theoretical and methodological comparison. The intersecting cross-disciplinary character of the study requires insights into many different fields of research, and much of the knowledge was developed along the way. After having conducted the research, I can see that it might have been rewarding to use an assessment literacy framework at the outset to guide the research and survey questions, rather than taking a more inductive approach. In sum, I could have limited my study to one school level to make the data more homogenous. To conclude, this single study cannot be representative of all contexts, but provides one contribution to a hopefully growing body of study on a fairly under-researched field.

Part 4 - Comparing contexts

Synthesis of results

Due to the meteoric rise of the minority student population, [...] and the importance of academic language use throughout the school, we have welcomed in an age where every teacher is now a language teacher. (Gottlieb, 2016, p. 1)

This study includes two different teaching contexts, sharing one important common denominator: the students in the subject matter courses in question are all language learners and are taught and assessed in a non-native language—an L2. In this section, the findings in Studies I and II are synthesized and discussed in terms of validity, equity and fairness in assessment among language learners. Thereafter, possible implications for language policy and language use in education are discussed while comparing the two contexts (CLIL and NAS) in relation to the overarching research questions:

- What are teachers' assessment beliefs and assessment practices regarding language learners? Do they differ depending on context, CLIL vs. NAS? If so, how do they differ?
- What are teachers' conceptions of language in relation to the subjects and policy? Are there common features between CLIL and NAS? If so, what are they?

Following a discussion of the main topics (educational assessment and language policy), the chapter concludes with possible implications for the teaching profession and educational organization. To summarize, the synthesis proposes possible directions for the future to strengthen validity, fairness and equity in educational assessment, before finishing with some suggestions for future research. However, before looking at the questions of interest, a brief comparison of the multilingual contexts (CLIL and NAS) introduces this part.

Comparing CLIL and NAS—language by design or by default?

The focus in the present study has been on two different language and content integrated learning contexts: one involving more or less homogenous classes where English is used as the medium of instruction among Swedish students, and the other representing heterogeneous classes with newly-arrived language minority students who are taught subject content in Swedish. The contexts could be defined as voluntary (CLIL) versus involuntary minorities (NAS), a distinction made by Ogbu (Cummins, 2017, p. 34), in that the first group chooses to migrate, or are preparing for future prospects in the global arena and the second has been forced to migrate (e.g., asylum seekers). CLIL is often considered beneficial for the learning of a second language where non-language content is used as a vehicle for promoting L2 proficiency (Genesee & Lindholm-Leary, 2013). The pedagogy used is often referred to as *bilingual education*, where the language of instruction (L2) and the students' L1 are perceived of as enjoying equal status. The other context (NAS with a migrant background, most of whom are refugees) represents minority groups with varying school backgrounds starting education in the dominant language (in this case Swedish) in the host community, a different type of bilingual education, or not, as discussed in the following. The character of the CLIL and NAS pedagogies in the present study will be juxtaposed and compared in a section below.

Lambert (1955) made a distinction between *additive* versus *subtractive bilingualism* based on learning conditions; the acquisition of two languages in a balanced manner is referred to as additive bilingualism (Cummins, 2000; Lasagabaster, 2009; Thomas & Collier, 2002). The additive perspective means that students add a language to the languages they already know. In the Swedish CLIL context, the additive perspective instead means adding a more advanced academic English proficiency to an already advanced level of L2 English (cf. Parts 1 and 2). This is contrasted with subtractive bilingualism, which refers to the education of learners with migrant backgrounds, for whom learning the second language implicitly or explicitly implies replacing the first language with the dominant language of instruction (Banks, 2009; Cummins, 2009). In Sweden, mother tongue instruction is stipulated in the Education Act (2010 Ch.10 §7) and has been in place since the 1960s (cf. Part 1, p.50), with one of the incentives being to avoid subtractive bilingualism. However, the fact that students have a right to participate in mother tongue instruction does not necessarily imply that all students do: only 59% of students entitled to mother tongue instruction received it in the 2018/19 school year (National Agency for Education, 2019a).

Øzerk (1998) developed an illustration of additive versus subtractive language development that inspired Figure 4.1 below.

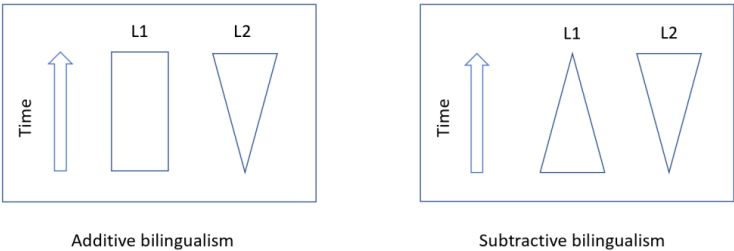


Figure 4.1 Illustration of additive versus subtractive bilingualism
Adapted from Øzerk (1998, p. 150f)

In additive language development, illustrated in the left-hand box above, both the L1 and the L2 reach similar (if not equal) status over time. In subtractive bilingualism, as illustrated in the box to the right, the L1 skills supposedly diminish as the L2 proficiency increases. In relation to the present study, I argue that the situation is more complex, since different language registers necessarily need to be taken into account, as shown in Figure 4.2 below. Cummins (1979, 2008), who developed a distinction between different language registers (basic interpersonal communicative skills and cognitive academic language proficiency), argues that it is problematic to incorporate all aspects of language use or performance into just one dimension of general or global language proficiency. In Parts 1 and 2 of this study, the concepts of content obligatory (CO) language and content compatible (CC) language were mentioned as being two important features of academic language. It has been argued that mastering only the CO language in a subject limits both the students in their expressiveness and the overall quality of their language skills, even though it may suffice to fulfill the course requirements (Köppe & Nijsten, 2017, 2012). Instead, it is important to turn all teachers into language teachers (Gottlieb, 2016; Schleppegrell, 2004, 2005, 2011). Ball et al. (2016) make a distinction between concepts, procedures (skills) and language, meaning that language goes beyond the study of subject concepts and subject skills. Hence, language includes the “other” academic language found in the disciplinary genre and the general academic language (Schleppegrell, 2005).

In the present context the categorization additive and subtractive bilingualism is not that straightforward. The CLIL students’ language proficiency at the outset given the extramural/out of school exposure to English (Sundqvist & Sylén, 2016; Sylén, 2005) alters the scenario, and as is the case for NAS: the age factor of late arriving students, the potential impact of mother tongue instruction and as mentioned above, exposure to different registers must be taken into account. According to research on CLIL in Sweden (Sylén, 2019; Yoxsimer Paulsrud, 2014), students who are confident English L2 users are particularly well versed in everyday language, which Cummins refers to as basic interpersonal communicative skills (BICS). However, the students’ proficiency level in the academic registers, expressed both as subject-specific literacy and general language of schooling, or cognitive academic language proficiency, is not necessarily as advanced at the outset.

Figure 4.2 below attempts to illustrate possible language development in the two contexts, CLIL and NAS. The size of the bars and cones is not empirically based but is meant to serve as an indicator of a need to pinpoint responsibility for students’ progress in different language proficiencies or registers.

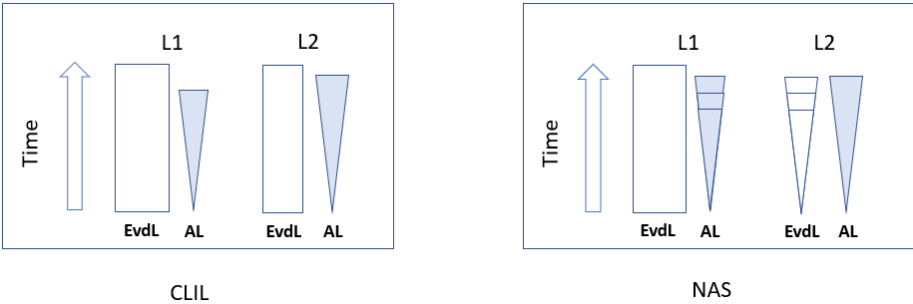


Figure 4.2 Illustration of additive bilingualism in different language registers (CLIL vs. NAS)

The white bars and cone represent the everyday language register (EvdL), and the blue cones the academic language (AL) registers. At the outset (at the bottom in each square), students have a certain proficiency in their L1 in both contexts. Since students in the present context are aged 13 years and up, it is assumed that they have attained a certain level in their L1 that will remain, provided they can use it with other speakers outside of school. When it comes to the academic registers in the first language, studies in the CLIL context have shown that students do not progress as much (Sylvén, 2019), and as mentioned earlier, subject matter teachers are found not to teach the general academic language in English, but focus mostly on concepts (Reierstam, 2015; Yoxsimer Paulsrud, 2014). Since English is the target language in CLIL, where bilingualism in Swedish and English is not explicitly pursued, students risk losing conceptual knowledge in Swedish, causing a certain domain loss in the academic registers in both Swedish and English. Among NAS, academic language proficiency depends on their previous schooling history (the blue cone to the left in the right-hand square) and their proficiency in the everyday Swedish register depends on their level of exposure outside school (the white cone to the right).

Teachers in the present study testify that they sense a certain lack of everyday language among the NAS, while the students are familiar with subject-specific concepts they have been taught in class. In both boxes (CLIL and NAS), the cones to the right have not reached the top since teachers in both contexts in the present study acknowledge that they focus mostly on concepts and less on general academic literacy in the subjects. I argue that for language to be learned “by design” and not “by default,” there must be a conscious language learning agenda that includes more than just conceptual knowledge.

Now, using the illustration in Figure 4.2, let us consider the following: What if...

- CLIL students experience less language development in the academic registers in Swedish, their L1, since instruction is targeting only English? (left-hand box above)
- CLIL students experience a lack of development in academic English, their L2, due to an unbalanced focus on concepts rather than general academic language? (left-hand box)
- NAS students experience less language development in the academic registers in their L1, if instruction is targeting only Swedish in the subject content courses and mother tongue instruction is not in place, at least not with “subject matter experts”? (right-hand box)
- NAS experience a lack of development in academic Swedish, their L2, due to an unbalanced focus on concepts rather than general academic language, similarly to CLIL? (right-hand box)
- NAS are at a disadvantage when acquiring disciplinary language due to poor everyday language in Swedish (L2)? (right-hand box)

More questions could be asked, but they are beyond the scope of this study. In the figure above I have chosen to use only the concept of additive bilingualism, but to varying degrees rather than subtractive, since subtractive presupposes a language level at the outset that is diminishing. Again, it is not the aim of this study to define the students’ language proficiency, whether at the outset or after a course. In the subject disciplinary registers, it is difficult to know even with native speakers how much they knew before starting a course, at least in the Content-Obligatory language (e.g., physics concepts or definitions of legal processes in social studies). What is of interest are the possible consequences for assessment and educational language policy in subject matter instruction, as discussed in the next section.

Comparing CLIL and NAS reveals varying conditions that apply as regards the students' language proficiency in the target language at the outset, the language diversity in the classrooms, students' exposure to the language outside of class, the teachers' L1 and thus proficiency in the language of instruction, to mention just a few (for more, see Table 1.10, Part 1). The teachers' language proficiency has not been explicitly discussed in this section but will be addressed in relation to teachers' professional development below. Regardless of differences in context, I argue that different immersion or language integration contexts can inform one another, whether it be what Gallagher and Leahy (2014) refer to as immersion by design (e.g., CLIL) or immersion by default (NAS or language minority students). As discussed in the section on language policy below, immersion might be a more appropriate term than bilingual instruction in the contexts of this study, as it implies more incidental language learning. I further argue that teachers in both contexts need to particularly consider the content and language integrated learning conditions in assessment for reasons of validity, equity and fairness; I delve into this aspect further below.

Implications for assessment

The appropriate development and use of assessments are essential requirements for responsible professional practice in educational testing and measurement. (Camara, 2015)

Little research has been conducted in the field of teachers' assessment of language learners' subject content knowledge, especially when described as "summative assessment." When assessment is mentioned in relation to language learners it is often in terms of *formative assessment*, *assessment for learning* or *assessment as learning*. One possible explanation, other than the perceived positive effects of the feed-forward potential produced by good formative practices, could be the identification of language learners as an underprivileged minority group, one at risk of not "measuring up" to the predefined knowledge requirements. I note along with Llinares et al. (2012), however, that the key to ensuring that students meet content and language learning goals is the principled and planned use of formative feedback, and you cannot have one without the other (i.e., both formative and summative assessment are needed). In order to provide students with equal access to the content in the subjects and equitable opportunities in assessment, teachers need to be familiar with subject-specific and academic patterns of language and engage students in metalanguage to talk about language (Schleppegrell, 2020). As teachers expressed in both studies, truly assessing knowledge and not language is difficult. The question is rather whether this is even possible.

As much as research in the migrant context deals with matters of identity, as mentioned previously, and talks about avoiding referring to deficient language or inability to reach standards, I think the teachers in this study touch on an important fairness issue. As one of the teachers notes, the students, in this case NAS, need to be taught how to use language lest they risk being continually characterized as "lower class." One issue in relation to late-arriving newcomers relates to whether national standards and learning goals can be considered reasonable or even attainable. The syllabi express knowledge requirements that presuppose equal and equitable groundwork. Researchers' and teachers' experiences, as exemplified in the present study, clearly point to the impact of diverging backgrounds and discrepancies in conditions among and between students of various backgrounds. As stated in the introduction, this represents a key issue when inequalities across students and groups are increasing (OECD, 2019a).

In a CLIL context, the issue is not so much a fear that students will not be proficient language users, but rather acknowledging that the goal to learn a language as a bonus, in this case English, must imply explicit language learning goals in the content syllabi together with the conscious teaching of the different registers and specific components of academic disciplinary language. Even if the students appear to be fluent in English, familiarity with the subjects' disciplinary language cannot be presumed, as mentioned in the previous section. Overlooking the construct of multilingualism is likely to result in tests of limited evidence to validity according to Shohamy (2011, p. 419) and therefore the students' language background and the intended language learning goals must be considered when aligning assessment and instruction.

Assessment is about knowing what the students know. For assessments to be valid and fair, the students must be given the best possible opportunities to show their knowledge and skills, together with accurate guidance on how to get there. An important aspect of validity in assessment is to ensure alignment between the target domain and the assessed domain (i.e., that the construct of assessment represent the same as what students have been prepared for). If students have not been given the chance to practice

the skills needed to perform on the assessment, this represents a threat to valid use (Gipps, 1994; Kane, 2010; Stobart, 2005). An important starting point is therefore to consider the students' background and performance level, including their language skills, in order to know how to help them reach the goal and perform their best. Factors that may inhibit students from performing their best have been identified as threats to the valid use of assessments (Crooks et al., 1996). The administration of assessments risks making it infeasible or unfair to students who are not native speakers of the language of assessment. To address this, different adjustments are suggested to accommodate the students' special needs, including alternative assessment formats, extended time, permission to use the strongest language and use of dictionaries or translation apps to understand the questions, to mention some of the most common (see Part 1, p. 24f). Accommodations are often perceived as necessary for reasons of equity (to "level the field") as a means to remove construct-irrelevant barriers so that the test format or the way the test is administered does not prevent students from demonstrating their knowledge, ability or skills. However, the use of accommodations is not without controversy, as noted by Sireci et al. (2003, p. 3):

At least two questions fuel the debate on the value of test accommodations. One question is "Do the test scores that come from nonstandard test administrations have the same meaning as test scores resulting from standard administrations?" A related question is "Do current test accommodations lead to more valid test score interpretations for certain groups of students?"

In the quote above, Sireci et al. point to issues that relate to the interpretation of test scores following the use of accommodations, asking whether the assessment can be considered to measure the same thing and if the chosen accommodations in fact help the students. For the same reason, accommodations are controversial if in fact everyone would benefit from the accommodation. Instead, this leads to an unfair advantage for students who receive them, meaning that scores are invalidly inflated, according to Sireci et al. (2016). Research has found two accommodations that seem to improve the performance of students with disabilities: extended time and oral accommodations. Extended time, however, seems to favor all students (Sireci et al., 2016), and therefore different treatment of students regarding time allowed for a task may represent a threat to validity. Teachers' empirically-informed knowledge of which accommodations to use to provide the appropriate support is therefore of the utmost importance to fairness and validity in assessment.

Accommodations often deal with the administration and implementation of assessment, but fairness also includes the *what* of assessment. Cummins et al. (2006) mention that bilingual students profit from writing about issues that matter to them in order to consolidate aspects of the academic language they have been reading, but also express their identities through language. However, for reasons of fairness and the risk of bias in the assessment situation, questions that require test takers to respond to personal questions could be regarded as construct-irrelevant content. As such, questions could be emotionally charged, thereby giving rise to negative feelings, and considerations need to be made as to whether these should be avoided (Educational Testing Service, 2016). ETS provides an illustrative list that is not intended to be exhaustive, but to give an idea of topics in relation to factors such as ethnicity, gender and disabilities that could cause students to not perform their best. Under the bullet point called *context*, the ETS guidelines (2016, p. 60) state:

For construct-irrelevant contexts of reading passages and math problems, the information required to key the items correctly should either be common knowledge among the intended test takers or be available in the passage or problem. Contexts should not require direct experience that is unavailable to people with disabilities.

“Unavailable experience” could be found in what teachers in the present study mentioned in relation to lack of ordinary everyday experiences like biking, which was referred to in physics assignments, for instance. Teachers also referred to the impact of differences in what is discussed and not discussed around the kitchen table in the students’ homes. That the required information is available in the test or test item is an example of context-embeddedness and is exemplified in assessment samples in both Studies 1 and 2 as vocabulary to be used, rephrasing via the use of synonyms, selected response items, providing initial letters, or the use of pictures.

In order to compare the teachers’ assessment beliefs and assessment practices in CLIL and with NAS and find out *if* they differ and in that case *how* they differ, three areas of comparison will be in focus below: how the teachers perceive of the students, how they perceive of what it is they assess, and how they assess. In the following, a brief overview of key findings is provided, summarized in relation to the research question

- What are teachers’ assessment beliefs and assessment practices regarding language learners? Do they differ depending on context, CLIL vs. NAS? If so, how do they differ?

Teachers’ perceptions of the students

In Study I in the CLIL context, the teachers acknowledge that it sometimes is difficult to know whether undeveloped answers or misconceptions depend on the students’ poor content knowledge or their lack of language skills. Teachers explicitly share that they tell their students that the use of English most likely will have a negative impact on their grades. The teachers thus show that although the CLIL students are described as ambitious and high performing in general, they are perceived as “deficient” to a certain extent in the academic registers of English. The teachers do not make any detailed analyses of exactly where and how the challenges appear, other than noting that they are most prominent in lengthier constructed responses or essay questions, and that they are language related. They note that students are particularly fond of written assessment and prefer to write since this leaves them time to think; this, at least, is the teachers’ interpretation of the situation. The teachers perceive oral assignments to be more difficult in that sense. The students also seem to like written tests, according to the teachers, since this is what they are used to.

In Study II, the use of the term “varied” is recurrent in the teachers’ descriptions of the students and their abilities. The students’ previous educational history or lack thereof, their varying opportunities to integrate quickly and interact with peers with a different L1, preferably Swedish, together with their ambition to learn, are mentioned as factors that will either fuel or impede their progress. In the same vein as for CLIL, undeveloped language is considered to be an obstacle. However, the NAS teachers note that students are often better equipped in the language of schooling than in everyday language, since they are taught academic concepts but not exposed to as much Swedish outside of school.

Higher grades such as B and C are feasible in some cases for ambitious students who are integrated into regular classes at either the upper secondary level or in Grades 7–9, although for a majority it is a matter of attaining a passing grade. Several teachers claim that NAS need extended time and at least one extra year to pass. Most teachers report, as for CLIL, that it is their impression that students prefer written assessment, except one teacher who works very consistently with oral assessments. In her case, the students must be convinced that they need to learn how to write, and should give it a try, even if they use both oral and written assessment formats. In the survey in Study II, the teachers reported that short written assignments and prepared oral assignments are easier, though not easy, whereas oral discussions

in class and longer written assignments are difficult; this resembles the experiences reported by the CLIL teachers. In the survey, the NAS teachers also noted that the most difficult language registers involve subject-specific written genres and subject-specific skills. Subject-specific concepts and everyday language are perceived as easy. The paradox in relation to everyday language was discussed in Study II, where it was noted that although teachers in the interview report deficiencies in this regard, it is possible that they make a distinction between lack of everyday concepts versus the use of simple sentence structures in the absence of academic language.

What teachers assess

Teachers in both contexts are very clear about the intended learning outcomes as expressed in the syllabi. They acknowledge the accountability purpose of assessment and a need to assess both conceptual knowledge and the skills that are expressed in the aim, the core content and the knowledge requirements. Teachers in both contexts also identify a need to assess both lower-order and higher-order thinking skills. Although the subject matter teachers in both contexts state that they cannot and should not assess language, seeing it as not their responsibility, they express that it is difficult to assess the more advanced critical thinking skills required by CDFs such as *draw conclusions* or *analyze* (Dalton-Puffer, 2016), where more language is needed.

In Study I, both the CLIL and non-CLIL subject content teachers infer that there are written genres that the students need to master and that these are part of assessment, such as lab reports, historical narratives or analytical argumentative essays. In Study II, the focus is often on getting the students to pass by merely answering questions and eventually having them write more coherent texts. In Grades 7–9, the teachers argue that the NAS will need those skills in upper secondary, and therefore written skills should be pursued. However, due to the students' insufficient language the teachers report that they use assessment formats where content knowledge is demonstrated in short written answers, oral follow-up, or dialogic exchange inside or even outside of class.

When the teachers are asked in the survey in Study II what is important in relation to their subject, a majority state that the use of subject-specific skills (e.g., drawing conclusions and analyzing) is particularly important, which actually coincides with what the teachers consider particularly difficult for the NAS. However, the teachers report that providing examples of how to construct argumentative speech or modelling subject-specific texts is not done very frequently in class. Subject-specific concepts, on the other hand, are frequently targeted, which coincides with what they consider easier for the students. The question is whether this is perceived as easy based on student results, in which case it may mirror what they have been taught. Either way, validity and equity in assessment are threatened if the assessed domain includes features that have not been taught; in this case, subject-specific skills and written genres.

How teachers assess

Teachers in both contexts favor written tests for summative purposes for various reasons. In Study I in the CLIL context, it is believed the students prefer written tests; certain share this belief regarding the NAS as well. Even though the NAS's teachers advocate the use of varied assessment formats, and several prefer and advocate the oral format, they conclude that it is not feasible as a basis for assessment with so many students. Teachers in both studies use mostly self-developed tests, and several use national

tests when applicable. When asked about question and item types, the teachers in both studies refer to using a mix of selected-response (SR) and constructed-response (CR) formats, since it is argued that the former only assess basic conceptual knowledge at the E level, the lowest passing grade. However, the teachers state that questions requiring more elaborate and creative responses are needed to find a basis for higher-order thinking skills and higher grades. Some teachers mention the use of old national tests as models or using parts of tests for assessment. They acknowledge that it is difficult to come up with good quality multiple-choice questions. The mandatory early skills mapping provided by the NAE to assess the literacy and subject content skills of NAS is not used regularly by the content teachers when planning instruction and assessment.

Comparing the two contexts, it is interesting to note how teachers perceive of which language should be used in assessment. Swedish is clearly the target language in the NAS context, and English in the CLIL context, according to most of the teachers. However, for CLIL at the upper secondary level there are no policies or explicit guidelines concerning language use. For NAS, there are guidelines for the instruction of NAS, and the recommendation is to use the students’ strongest language, at least initially. However, a majority of the teachers in this study state that they use Swedish, and that there are no specific guidelines for which language to use in assessment.

In Part I, an assessment literacy model was presented, inspired by Abell and Sigel (2011). According to the model, teachers’ assessment literacy includes four questions, three of which are used to compare teachers in the two contexts in Table 4.1 below.

Table 4.1 Teachers’ self-reported assessment beliefs and practices in the two contexts (CLIL and NAS)

	CLIL	NAS
Why assess	<ul style="list-style-type: none"> Accountability: Performance-oriented 	<ul style="list-style-type: none"> Accountability: Performance- oriented Pedagogy: Guides student learning
What to assess	<ul style="list-style-type: none"> Knowledge requirements in syllabi Subject skills Not language Written genre 	<ul style="list-style-type: none"> Knowledge requirements in syllabi Subject skills Not language (?) Written/oral genre
How to assess	<ul style="list-style-type: none"> Written format Using English (Using Swedish) 	<ul style="list-style-type: none"> Written format Oral format Mixed formats Using Swedish (Using mix of languages)
Accommodations	<ul style="list-style-type: none"> Dictionary Extended time Oral follow-up (Providing cues in the tasks) 	<ul style="list-style-type: none"> Translation apps Extended time Oral follow-up (Wordlists in the tasks)

Accommodations could be included in the *how*, as they were above, but since this is an important issue in relation to assessment of language learners, they were separated in the table. In the column to the right, features that clearly differ are highlighted in bold type. Parentheses indicate that something was random, and not a general feature. The question mark after “Not language” indicates that some teachers acknowledged that language is part of the assessed domain, whereas teachers in Study I were clear that it is not (although they found that language sometimes interferes with assessment outcomes). The fourth question from the assessment literacy model in relation to the consequential validity of assessment—*and?*—is discussed at the end of this section.

Certain differences are traced in Studies I and II between subjects and school levels. The accountability purpose of assessment is more pronounced at the upper secondary level (Study I and part of Study II). The pedagogical purpose (guiding learning and as wash-back on instruction) is more clearly expressed in Study II. The focus on summative assessment was explicitly stated by the researcher in both studies. In Study II, however, it was clear that teachers were more concerned about the poor results of the students, and thus felt a need to let the students’ results and progress inform instruction. Whereas assessment of written genre was used in CLIL, it was not feasible for NAS at a basic level due to their lack of language. On the other hand, teachers of NAS stated that short written comments and questions were easier for NAS than making oral presentations. Thus, a clear distinction appeared between dialogic oral follow-up, which was used as an accommodation, and prepared oral presentations, which were perceived as more difficult for the students, particularly in the NAS context. Teachers in both contexts expressed concerns about the validity of oral follow-up, since they felt that they might be helping the students too much, which would have a negative effect on the legal certainty of the assessment, according to them. For similar reasons, they expressed concerns about assignments done at home.

The NAS teachers reported favoring mixed assessment formats to a greater extent; however, digital assessments did not seem to be commonly used. In Study I, where data was collected in 2013, a couple of the EFL teachers mentioned the use of digital exams and in Study II, a couple of the social science teachers mentioned using a platform for students to post their written assignments. One of the teachers in Study II posited that the use of computers would imply an extra strain on the NAS.

The same types of accommodations were found in both studies, although in Study I they were not used by all teachers and they were not perceived as accommodations, e.g. more time on task, the use of dictionaries and oral follow-up after an exam. As regards time, the CLIL teachers said that the use of English clearly meant that they had to cut down on course content, since there was not time to cover the intended course content. In a similar vein, the students needed more time to process due to the language, and therefore time was not seen as an accommodation, but rather a prerequisite. Some CLIL teachers allowed dictionaries and some said that the students could choose to use either Swedish or English in assessment, but instruction was in English. Other CLIL teachers said that English should be used in assessments. Most of the NAS teachers stated that Swedish should be used, but some said that the students could use their strongest language if there was a teacher to translate. This resembles the concerns Kaya (2017) raised about inequality in assessment, which were referred to in the introduction of Part 3. If students are given unequal opportunities depending on their L1 and access to an interpreter, this is an issue in relation to equity and fairness. One interesting note between Studies I and II is what I perceive of as a higher awareness of the impact of language among the teachers in the second study. Although subject content teachers in both contexts claimed not to assess language and focused mainly on subject concepts, teachers in Study II expressed more concerns about how to model language for the students. In Study II, several of the content teachers explicitly said that they need to be language teachers in a sense. A possible explanation for this may lie in the different contexts, or may be due to the fact

that awareness of language in all subjects has increased in education recently (e.g. Bjerregaard & Kindenberg, 2015; Cummins, 2017; Hägerfelth, 2011; Karlsson, 2019; Kindenberg & Wiksten, 2017; Kouns, 2014; National Agency for Education, 2012b, 2018e, 2012c; Josefin Nilsson, 2015; Nygård Larsson, 2011; Swedish National Agency for School Improvement, 2018).

In summary, it can be noted that in relation to the question “*and?*” and the consequences of assessment, several threats can be traced in relation to construct-irrelevant variance, since the required language is not consistently and deliberately taught. This does not mean that teachers do not try: they work hard to cover course content, meet the knowledge requirements, and perform well in terms of helping the students to the best of their understanding and ability in both contexts. The problems are rather to be found in inconsistent or non-existent policy, the unarticulated responsibility for academic disciplinary language, and what seems as a lack of appropriate or sufficient teacher training given the multilingual teaching context.

I echo what was expressed by Gallagher and Leahy (2014), that despite major differences, my research points to distinct similarities between teachers’ practices in the two contexts. I believe the underlying structures and norms that shape the education system as well as teachers’ disciplinary beliefs affect assessment similarly in both contexts. I therefore see a potential for cross-fertilization and transfer across subjects and between contexts.

Teachers in both contexts:

- are faithful to the syllabi in assessment.
- feel the need for accountability in assessment.
- lean more heavily on written assessments (whether they like it or not).
- state that they are *not* language teachers (although teachers in Study II less so).
- use oral follow-up as an accommodation.
- believe that the target language (i.e., English or Swedish) should be used.
- do not collaborate very much with colleagues.

In the next section, issues in relation to language policy are addressed, followed by a discussion of teachers’ professional development and education.

Implications for language policy.

“Even where there is a formal written language policy, its effect on language practices is neither guaranteed nor consistent” (Oliveira, 2005, p. 2153)

Language policy is directly linked to language ontology, ideology and use. Dichotomous views coexist today in explicit or hidden agendas in education and society at large. Shohamy (2006) posits that language is manipulated and turned into a stagnated and oppressive standard variety that is advocated by politicians, linguists and educationalists, a phenomenon unknown to the public, who have been indoctrinated into perceiving language as a correct and pure system. This view correlates with a disaggregated post-modern ontology, opposing named, established languages and fixed language developmental stages, and seeing language as dynamic and socially constructed. Translanguaging policy and practices stem from this perspective. On the other hand, using the terms *standard*, *norm*, *language registers*, *target language*, or even *mono-*, *bi-* and *multilingual* denotes an underlying essentialist view, where language is seen as a fixed system, which can be traced in my approach to the *what* and *how* of assessment.

According to Spolsky (2004), “language policy” equals all the language practices, beliefs and organizational preferences of a society. However, whether explicit or not, language policy will not necessarily be consistently applied, as Oliveira (2005) notes in the quote above. In the present study, inconsistent language policies can be perceived in both the CLIL and NAS contexts, as noted in the previous section. The teachers’ beliefs and practices reveal a certain confusion in education today relating to what can be described as contradictory messages in policy guidelines and local practices that demand the teaching and assessment of a specific target language (i.e., English or Swedish), while at the same time advocating translanguaging. Target language teaching and translanguaging represent two policy approaches stemming from different world views. The vacillation as to how to conceptualize bilingual or multilingual education (or perhaps translingual) is symptomatic of the current trans-/cross-international confusion surrounding the definition and status of language(s) in the post-modern era. Schissel et al. (2019, p. 375) note that there are many unresolved issues in terms of the conceptualization of multilingualism within the field of applied linguistics, which creates challenges for assessment due to monolingual standards. Below, findings in the two studies are outlined in relation to the second overarching research question:

- What are teachers’ conceptions of language in relation to the subject and policy? Are there common features between CLIL and NAS? If so, what are they?

Perceptions of language use and language pedagogy

When CLIL teachers are asked how they deal with language, all teachers state that they teach subject concepts in English, and some CLIL teachers explain that they have learned from colleagues who are language teachers how to use flashcards with terms in English on one side and Swedish on the other. In a similar vein, several NAS teachers testify that they have learned some useful methods from language teacher colleagues, or they have a dual certification themselves, often in the social sciences and SSL. As noted in Study I, some of the CLIL content teachers say that they let the students choose which language to use; if they want to use Swedish, they can, since they are not expected to assess the students’ language. Some of the NAS teachers in Study II also state that they allow mixing or translanguaging;

however, a majority infer that Swedish is the goal, even if they claim that there is no explicit language policy. Several of the NAS teachers note that NAS make faster progress if they are separated from peers with the same L1, and that it is better for them to be forced into using Swedish as fast as possible, which implies faster progress in the content courses as well, according to the teachers. Teachers’ relationships to language in the subject disciplines seems rather straightforward, but not necessarily coherent or compatible with language integration “ideals.” In CLIL, the teachers seem to be informed by an idea of “language by design,” where the use of English is purposeful and provides a “silver lining” for otherwise standard teaching. However, teaching is not designed to target academic language learning. In the questionnaire in Study I, two of the EFL teachers were the only ones who reported general academic language to be important in assessment. In the same vein, the NAS teachers stress the use of subject concepts and skills, but not general academic language.

Below, the language integration model that was presented in Part 1 is used to illustrate the tension teachers in this context experience between the language demands of curriculum and language/content in pedagogy. The purpose of the model is to give a visual representation of the spread when integrating teachers’ instruction of language learners in subject content courses using more or less monolingual subject syllabi.

The language integration matrix combines how language is taught in class on the horizontal axis with how language is perceived in relation to the subject syllabi on the vertical axis. The subject matter teachers from the interviews are represented by letters (one per individual): C for CLIL, NC for non-CLIL and T for teachers in the second study. The interviewed subject matter teachers from Studies I and II have been positioned according to how their self-reported practices fit in relation to orientation to language in curricula and in their teaching. Thirteen NAS teachers, four CLIL teachers and two non-CLIL teachers have been inserted. The table should not be interpreted as a plot diagram: the positions of the teachers are entirely based on my interpretation as a researcher and should be read as the teachers’ “most typical position.” The figure, as well as the placement of the teachers, is only meant to illustrate perceived tendencies, since boundaries between the squares are not fixed and the teachers sometimes give contradictory messages.

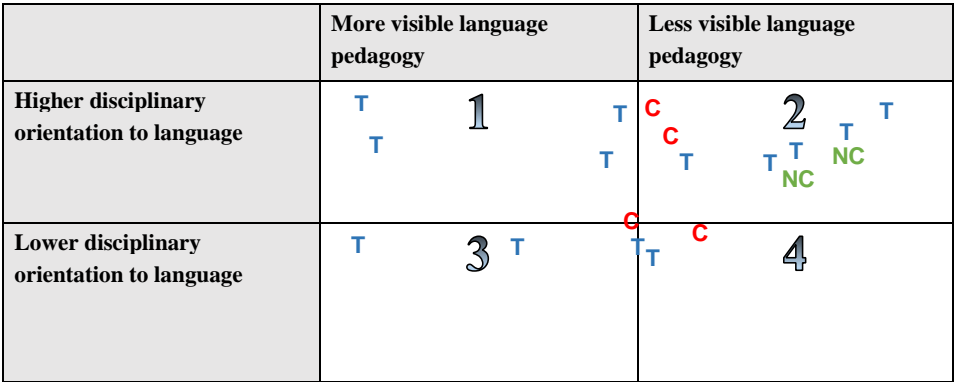


Figure 4.3 The interviewed teachers’ language integration profiles
 (blue T= NAS, red C = CLIL, green NC= non-CLIL)

It is interesting to note that there is a tendency for the subject content teachers to land mostly on the top half of the figure, which indicates a higher disciplinary orientation to language that I believe hinges on the importance of curricular goals expressed in syllabi and curricula. There is also a tendency to cluster to the right, with less visible language pedagogy and a focus mainly on concepts, which has been revealed in the findings of both studies. Whereas more visible language to the left implies more explicit language-oriented instruction and the prioritization of language over content, the less visible language pedagogy to the right implies incidental language learning and the prioritization of content over language. The latter attitude can be found in the beliefs of non-CLIL as well as CLIL teachers, although CLIL teachers are sometimes on the border with some explicit language instruction; for the most part, however, “explicit language” represents a focus on concepts.

Lower disciplinary orientation at the bottom half implies more decentralization and diversification, meaning a focus on general literacy rather than specific disciplinary language in situations where L1 use is not a problem. Content meaning is created in dialogue and translanguaging may be the norm in Square 4. None of this is seen consequently in any of the contexts, although some NAS teachers describe using dialogic speech events in class where language is created in use, which belongs more in the bottom half of the matrix. Two of the CLIL teachers—one in biology and one in history—claim that students can choose to use Swedish, but they (the teachers) reply in English. Already when summarizing the descriptions provided by Leung and Morton (2016), but even more so when applying the profiles, it became clear that Profile 3 appeared slightly contradictory. With a lower disciplinary orientation to language as seen at the bottom half of the matrix, it becomes less clear what is to be taught, yet Profile 3 implies more visible language pedagogy. A “translanguaging norm,” with a diverse language view thus seems more natural in combination with less visible language pedagogy, as can be seen in Profile 4.

As seen in Figure 4.3, teachers are all over the matrix, but there are slightly fewer in the bottom half, which represent decentralization and diversification, as in translanguaging practices. Four of the teachers are a little “borderline,” displaying some translanguaging approaches, but the explicit demands of the course requirements naturally push toward a higher disciplinary orientation to language, and thus rather inhibit total adherence.

To provide background to illustrate the profiles, descriptions and examples in relation to each, Profiles 1–4 are provided below.

Profile 1:

Explicit language-oriented instruction Higher disciplinary orientation to language Centralization and unification
<ul style="list-style-type: none"> • Clear descriptions of linguistic goals. • Specific stages of acquisition. • Language before content. • Focus on explicit subject literacies and skills. • Turn function verbs into learning objects. Can-do statements as in the CEFR. • Scaffolding to remedy the use of an L2. • Performance oriented =What is absent/deficient?

- T9 For example, when it says to be able to draw conclusions, yes, but what is a conclusion? And we might look at examples in class, how to see relationships between different events, for instance, what can you look for, what do you need to think about when you develop an argument, support sentences on the one hand, and on the other, what can help the discussion forward, connector words and so forth
- R Do you do this together in class?
- T9 Yes, we do it together, but we feel it benefits everyone
- T9 I use a lot of assignments to practice, before a test we have actually worked on every single question, and then they get to show [their knowledge in the assessment situation]
- R Do you think it is rewarding to work like this [with language]?
- T13 Yes, it is very rewarding I must say, simply because you develop your subject in a different way, it becomes very didactic, [reflecting on] how to teach. [...] I understand everything [in my subject] but it is easier to work with language development to help the students understand too, to read and write, or whatever it is
- R Would you say that you become more of a language teacher?
- T13 Yes, a little maybe, well, yes, if you work with language development you do what a language teacher does in a sense, absolutely, but I think all teachers should do it in that case, one might say that everyone needs to be some kind of language teacher

Example from the CLIL context:

C3/Hi/C: On the whole I work a lot with the writing process and try to target that it is the tool of the historian; history is basically the literary genre.

In the examples above, language comes before content with a focus on subject literacies and skills, and the students receive scaffolding through support sentences, for instance. The example from the CLIL context is representative of more visible language policy with a clear focus on subject-specific literacy, implying the top half of the matrix. At the same time, the teacher later states that the students can use Swedish if they want, which represents translanguaging practices. Again, the coherency between language learning goals and language practices is not straightforward in the absence of policy.

Profile 2:

Incidental language learning Higher disciplinary orientation to language Centralization and unification
<ul style="list-style-type: none"> • Natural acquisition view of L2 (immersion). • Already existing abilities can flourish and grow. • Content before language, language an additional benefit. • Focus on discipline-specific concepts and competences. • Language a tool for participation. • Competence oriented = What is present?

- R [Do you help them with] the other language that they need to make sentences, and answer questions. Do you work with that too?
- T1 [...] I don't work so much on that, but of course I comment and help sometimes, [...]
- R So you give them some hints, make corrections?
- T1 Yes, I do, that is part of working with this group, to work on the language.
[...] but that is something you should ask the Swedish language teachers about; I try to teach about society, and the world we are in, that is my job.
- T11 I often think that a film gives a picture of what we are to learn, this is what we need to understand, but you need to present [key] concepts in the film, since they don't understand those, you need to set the focus, or tell them this is what we are going to concentrate on. But then it is a matter of accommodations again, you don't have to pick 50 questions, you may choose 10 relevant questions, and understand those, and that covers the key concepts at the same time as you connect this to the film. [...] My experience is that it "ties in" [with their previous experiences] and then you can refer to their home countries

Example from the CLIL context:

B5/Bi/C: I don't know how much I work on that [language], more than in the classroom where there is a lot of talk. [...] It is based on these key concepts, but I put them in sentences, to create a story, and in that case, it is my way of talking which is either a help or insufficient help. [...] But I have not experienced that this should be a problem.

In the examples above, there is a clear focus on discipline-specific concepts, but content is prioritized over language; language is more of an additional benefit. The CLIL teacher mentions providing some context as scaffolding and reflects on whether this may be of help or not. However, she does not express that the general academic language is to be targeted in instruction, since it is her impression that it does not represent a problem.

Profile 3:

Explicit language learning Lower disciplinary orientation to language Decentralization and diversification
<ul style="list-style-type: none"> • Focus on general literacy norms, less attached to specific disciplines. • More or less functional. • No explicit stages of acquisition. • Lessons seen as speech events using function verbs, sociocultural perspective. • Individual choice and agency. • Teachers mediating and making content accessible.

- T3 Even if the schedule separates civics or Swedish it becomes like a big civics/Swedish topic, because language is present all the time. [...] so that is the point with me doing both the subjects, even if I am not certified to teach the social sciences, I can do language all the time, so there is a lot of focus on the language.
[...]
[We] always start with the concepts, I post them on the walls, to give them the basics, the foundation, sort of. Then we do a lot of things TOGETHER, read the texts, talk a lot, explain, they get to discuss in groups a lot. We do a lot of talking and processing together. [...] but I think the doing together is important, and to show by example, not only the language but the skills, because according to Swedish curricula you must be prepared to draw conclusions and analyze, and these students are not so used to this. They are used to factual recall and learning by heart, so this takes a lot of practice.
- T4 I like dialogic teaching, but I am also quite traditional during many weeks of a school year, but my lectures consist less and less of pre-made PowerPoint presentations with pre-created answers. Instead, I use pictures and in dialogue with the students I use the board to write down things they come up with and use that for the discussion. When they are involved in the discussion, they can practice the verbal skills so they can answer a question in a test.

In these examples, lessons are seen as speech events, representing a sociocultural perspective with teachers mediating and making content accessible, taking the students' needs as point of departure. Yet the example from T3 could also fit in Profile 1 in the sense of turning function verbs into learning objects, since both Profiles 1 and 3 represent more visible language pedagogy. In the CLIL context, no clear example can be found to represent this profile. The example used as an illustration for Profile 1 is a little borderline and could be used here as well, but the teacher makes very explicit comments about the literary genre in the specific subject discipline (history), implying a higher disciplinary orientation to language.

Profile 4:

Incidental language learning Lower disciplinary orientation to language Decentralization and diversification
<ul style="list-style-type: none"> • No strong separation or classification of linguistic varieties (L1 or L2) or subject-specific registers. • Language competence is multifaceted, multilingual and multimodal. • Expand communicative repertoires to be able to participate. • Translanguaging the norm. • L1 use not a problem. • Content meaning = created in use, dialogue, not in words.

- T3 They can use their own language when they are answering questions in a test. If they don't have the words they can write in Arabic. Most of them speak Arabic, and that is an advantage since I have a colleague who speaks Arabic and he can help me translate. Those who know English can use English.
- T3 In a test, I often get a lot of mixed languages, some in Swedish, a word of Arabic here and there, some more in Swedish...
- R Sometimes you hear translanguaging?

- T3 Exactly! And they can draw pictures, I try to make them use all they can to explain what they think.
- T12 The first thing we do at the beginning of every module is a mind map, a conceptual map; “what do you know about the topic?” [...] Then we translate together, to work with translanguaging. Then we do it together on the board [...] and then we read the text together, they repeat after me, and then depending on their level they get different assignments. [...] After that they get to copy the text some are not illiterate [...] and they can translate the text to their language.

Example from the CLIL context:

B3/Bi/C: I don't require that they have to know specific concepts when it comes to the kidney; renal pelvis for example. They probably know “kidney”, but if they don't remember renal pelvis, they can write it in Swedish. [...] I don't require that they have to know specific words, that is not the main point, they need to know relationships; how does the kidney function—describe [...] If they don't know all the words in English, then it works just as well in Swedish. They need to know how to describe what happens.

B3/Bi/C: It is better they that they say something, even if it is in Swenglish or Swedish, than them not saying anything. It can be a bit tricky, but it is not the language that is the main thing, it is still biology that is the main thing, so to speak.

In these examples, teachers advocate translanguaging more or less explicitly. Use of L1 is not a problem and content meaning are created in dialogue. As we can see, T3 appeared in Profile 3 as well, and could have been in Profile 1, since several of the features are combined in the same practice. Teachers' alternative practices should not be considered problematic as long as there is a coherent and aligned purpose leading up to clear learning goals. Fixed positions would be more problematic if this meant not providing necessary access to what is to be assessed later, or not being flexible in meeting the students' needs. Translanguaging, for instance, can be used as an interim strategy as long as the target language is taught as well. Hence, this would mean shifting practices between Squares 4 and 1, since 3 is not really an option in Swedish schools as standards are very clearly articulated in subject syllabi.

The threat to validity in assessment relates to whether decisions are substantively unfair, whether scores mean different things, and whether opportunities and access differ across contexts. Differences occur when no clear policy exists, as the findings in this study have revealed as regards teachers' perceptions about which language should be used, varying pedagogy and unequal use of accommodations and access to translators for NAS. The consequences of different assessment practices are discussed below.

In a Swedish context with very clear language and content integration in the intended learning outcomes of the knowledge requirements, a shift to the left in the model would be desirable (see Figure 4.3). This would align instruction with the learning goals and prepare students to use cognitive discourse functions and skills, modelling disciplinary genres. In consideration of the importance of the knowledge requirements in syllabi, Square 1 would be targeted, adding explicit language goals to the ILOs, which are more or less implicit at present. If a translanguaging norm is the goal, and not just a strategy for the initial phases (as in a lower disciplinary orientation to language), this would necessitate less explicit language integrated knowledge requirements, allowing more freedom of choice. A “square” model such as the one shown above is a blunt tool, and flexibility is not built into the figure. A cross-fertilization between clear language goals (at the top) and dialogic language learning (at the bottom) would seem

most desirable; indeed, many teachers are targeting this in the present study, many of whom, but not all, have a dual teacher identity (which here typically implies teaching both the social sciences and SSL or EFL).

In relation to the definition of the target language (TL) and intended learning outcome (ILO), instruction and assessment need to be aligned. Whether aligned or not, language use will impact the consequential “And?” at the end, both in relation to what has been taught/learned (and not), and what are perceived as future language needs, regardless of students’ L1. (See Figure 4.4 below.)

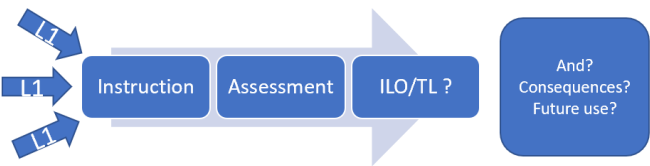


Figure 4.4 Alignment of languages in education with intended learning outcome/target language

Constructive alignment, according to Biggs (2003), means aligning instruction and assessment with the ILOs. As explained above, this includes a TL, which is expressed as curricular goals and thus means both disciplinary language and the language that is needed to use the knowledge to communicate in society, in future education, and in the assessment situation at school, which is meant to provide a “receipt” for the attainment of certain knowledge. From a translinguaging perspective, this allows for many possible languages, but questions need to be resolved in relation to how functional this is in the end, and in which contexts it can be used, if the result is a continued mixing of languages.

If on the other hand, translinguaging can be perceived as a strategy. By using multiple languages and language registers, mixing both L1 and L2 and academic language with everyday language as an interim solution, it can help creating an *interlanguage* on the path toward the target language, as illustrated in Figure 4.5 below.

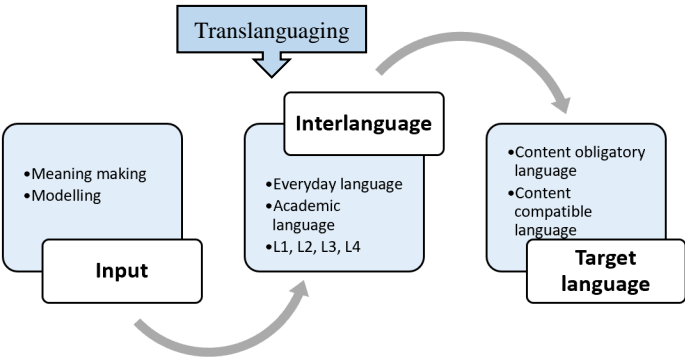


Figure 4.5 Language processing stages towards the target language of schooling

The figure tries to capture the language learning process, starting from new input, where the teachers and students are working on sense-making and hopefully working on modelling language. Schleppegrell (2005, p. 47) states that teachers need to engage students in discussion about language that raises their awareness about the way knowledge is constructed in language in different subjects, something that is neglected in many classrooms. At the same time, Gibbons (2018) speaks about producing hybrid discourses via register meshing, shifting between familiar and technical language and written and spoken language. Here, the individual's different languages (L1, L2, etc.) are added. The inter-language forms or stages in the middle are considered "temporary" (Gottlieb, 2016; Olander & Ingerman, 2011), since the goal is to know a subject as described in the national curricula, where all students' right to know enough academic Swedish is described. Eventually, the target language materializes, specific registers are recognized, separated and appropriately used. In Figure 4.5, the academic registers are labelled *content obligatory language* versus *content compatible language* (see Table 1.2, p. 17); in other words, subject-specific versus general academic language.

Translanguaging has been purported to boost motivation and bridge the gap for late newcomer students in desperate need of speeding up the language learning process in subject content courses. Issues occur when assessment is at stake. In the absence of clear policy guidelines for translanguaging and assessment, some important questions must be addressed: Which language(s) should be used? When should a target language be expected? Can a non-target language be as valid for showing understanding and knowledge of subject content? How should "insufficient" language be considered in assessment in the meantime? And how does it affect students' abilities to achieve higher grades? Can assessment be done without language, or at least require less language? Following her study on translanguaging as a learning resource in the science classroom, Karlsson (2019) concludes that concepts and words in both the L1 and the TL (Swedish) tend to be simplified, which has severe consequences for the continuity of science learning in classroom activities. The usefulness and practices of translanguaging thus need to be thoroughly investigated, especially in relation to assessment, if it will contribute to valid and fair assessment outcomes.

To conclude this section on teachers' perception of language use and language pedagogy in the CLIL and NAS contexts, the figure below provides an overview:

Intended (Planned)			Implemented (Enacted)		Attained (Learned)	
Policy What/how			Beliefs and practice What/How		Student learning outcome What/how	
	Grades 7-9	Grades 10-12	Grades 7-12		Grades 7-12	
CLIL	50% Swedish	100% English?	CLIL	English a bonus/ English the goal/ En + Swe the goal	CLIL	= Monolingual? = Bilingual? = Multilingual? = Translingual?
NAS	100% Swedish Trans- languaging?	100% Swedish Trans- languaging?				
CLIL	Functional academic English and Swedish	Functional academic English? Swedish?	NAS	Swedish the goal/ Swedish + L1 the goal	NAS	= Monolingual? = Bilingual? = Multilingual? = Translingual?
NAS	Functional academic Swedish	Functional academic Swedish				
CLIL/NAS	Knowledge requirements	Knowledge requirements	CLIL/NAS	Knowledge requirements	CLIL/NAS	Knowledge requirements?

Figure 4.6 Comparative curriculum assessment in Swedish CLIL and NAS contexts

Intended curriculum: The knowledge requirements are the same in both the CLIL and NAS contexts, but since language policy actually differ, they are divided into different school levels on the vertical level. In CLIL, no explicit policy exists in Grades 10–12 as regards which language to use, and therefore teachers can choose to use only English; up to Grade 9 in compulsory school, however, no more than 50% of the instruction can be in another language, i.e. in English according to the Ordinance for compulsory school (SFS 2011:185, Ch.9 §18-19). This may suggest an intent that half of the instruction in every subject should be in Swedish, however this is not necessarily how it is done; finding bilingual subject content teachers can be very challenging to say the least. In certain schools with an international profile that were not part of the present study, half of the courses are therefore done in only English, as stated on this website: “All courses, with the exception of Swedish, Civics A, Law and Modern Languages, are taught in English” (IES, 2012). This may in turn raise questions about domain loss in Swedish and the status of other languages if schools were to offer the same kind of bilingual education, but not in English. A goal of the Swedish curriculum for compulsory school is to make all students functional in academic Swedish, but English is also stated to play a central role in the Swedish school. Toth (2018, p. 81 f) points to the hegemonic dominance of English in Swedish bilingual forms of education, which is supported by national education policies and local teacher recruitment policies that privilege native English speakers.

Implemented curriculum: The implementation of the intended goals refers to the actual teaching and learning activities taking place in schools through interaction between learners and teachers as well as among learners (i.e., how the intended curriculum is translated into practice and actually delivered). This can also be defined as the “enacted curriculum.” Whereas Swedish or Swedish and a mother tongue/L1 is the goal among NAS, it is less clear in the CLIL context whether English is perceived as a bonus or as the actual goal. Although good CLIL practice is defined as bilingual education, where both L1 and English should be pursued, this is not typically the case in Sweden.

Attained curriculum: The attained or learned curriculum indicates the knowledge, understanding, skills and attitudes that learners actually acquire as a result of teaching and learning, and is assessed through

different means and/or demonstrated in practice. It may differ from the intended and the implemented curriculum. Referring back to the initial discussion in this part, and how to define the two contexts, the literature would suggest that equal status bilingual education should be the goal in CLIL as well as among NAS, providing conditions for language development in both the L1 and the TL, an L2. In Sweden, both contexts could rather be characterized as *immersion programs*, where a great deal of focus is on the target language. The CLIL context might even be characterized as English monolingual. Even though Swedish is the focus in the subject matter courses for NAS, legislation and policy explicitly state that they should have support from study counselors in their L1 (National Agency for Education, 2016). The question marks to the right in Figure 4.6 after attained language proficiency are meant to illustrate uncertainty in what is achieved as regards language skills. The big question mark to the bottom right signals uncertainty whether the knowledge requirements are attained, and if so, which requirements, and what are the consequences for the individual and society?

From an educational equity perspective, the bi- multi-, pluri- and translingual teaching contexts need to be critically examined in order to possess language learning practices by design rather than by default, and to achieve a corresponding alignment with valid and fair assessment.

Linguistic hierarchy in Sweden

Hyltenstam et al. (2012) maintain that all of the languages present in society represent important assets, which tends to be forgotten in a political discourse in which multilingualism is perceived to be a problem. The question is how to find a resolution between, on the one hand, advocating and striving for language competences for NAS in the “host country” and not risking domain loss for Swedish in CLIL, while on the other hand safeguarding the development of other language competencies in society. Students with a broad linguistic repertoire should be encouraged to keep and use their languages, put them on their résumés, and feel proud, since all languages are assets. There are many languages that are declining in Sweden, including foreign languages taught in Swedish schools such as French and German. The figure below illustrates the current language hierarchy, a figure borrowed from Hult (2012).



(A visual representation based on Josephson, 2004, p. 128)

Figure 4.7 *The Linguistic hierarchy of Sweden*
(from Hult, 2012, p. 242)

In Sweden, one-fourth of students have a foreign background and thus have a first language other than Swedish. At the same time, the proportion of Swedish students choosing schools where English is used as a medium of instruction is increasing. Recent reports from the OECD (2017, 2019a) call attention to issues of segregation in Swedish schools. The findings in the present study can be said to illustrate two separate categories on the spectrum. They represent different conditions yet are guided by the same curricular goals, where Swedish is synonymous with access to education and English with access to the world.

Before moving on to suggestions and implications for organization and teacher education, quotes from three of the teachers in Studies I and II will end this part on language and assessment in multilingual education. Teachers in Swedish schools obviously cannot know all the 165 languages represented among NAS in Sweden, or the 50 languages that sometimes end up in the same school (SOU 2017:54), but the subject matter teachers in both Studies I and II are trying to figure out how to handle language in their courses. Some are pondering whether it would have been easier if they had known the students' languages:

T7 [B]ut of course, in the best of worlds, it would be good [...] if we could have teachers with many different languages, or if you knew the languages yourself, then it would have been much easier.

...or using a strategy to learn from the students, while at the same time making them use Swedish as a lingua franca:

T11 A colleague, she made groups with three of four different languages in each, [...] and the purpose was to speak Swedish, and that is good, if you turn things around, we cannot talk anything but Swedish and yet we are so different so we can learn different things from each other.

...or reason as the CLIL teacher who believes that language should not be assessed regardless of context:

C3Hi/C If I am unsure [what a student means] I have a discussion; "I understand what you mean, so can you think of this next time". But the English proficiency cannot be part of an examination just as Swedish should not be part of assessment in history.

In summary, language is to be used, it is key to communicating and opening new worlds, but it should not be assessed for fear of unfair treatment. Language and assessment seem to represent two stumbling blocks in multilingual education that need to be addressed. Below, I offer some suggestion as to how this can be done.

Implications for organization and the teaching profession

Responsibility constitutes a factor that is not always explicitly stated in relation to teacher beliefs, but instead in laws and regulations. It is associated with the students' equal rights to education, and ultimately each school or educational provider can be held accountable. The question is: Who is responsible for fostering generic "enabling" transversal skills, as in academic language and literacy? Is it the content teachers' responsibility as subject matter experts to model academic language for the students, as in this example from a CLIL teacher in Study I?

B5/Bi/C I don't know how much I work on that [modelling a disciplinary genre], more than in the classroom where there is a lot of talking. [...] It is based on these key concepts, but I put them in sentences, to create a story, and in that case, it is my way of talking that is either a help or insufficient help. [...] But I have not experienced that this should be a problem.

...or some language teacher colleague's, as noted by the European Council?

"Although language as subject should not be seen as a "service" subject providing the necessary language teaching that is then used in other subjects, it does have a special role to play in relation to language education" (Beacco et al., 2016, p. 66).

Is it up to the Swedish language teachers, as expressed by one of the teachers in Study II?

T4 [There is] a lot of pressure on the Swedish language teachers, they are the ticket to upper secondary, pretty much.

Or, indeed, is academic language learning the students' own responsibility? Policy documents express the need for students to develop language skills to become proficient users of the national language in various segments in society, but depending on how language is viewed, it becomes more or less visible in pedagogy and more or less explicit in the disciplines, as seen above. Teachers with experience from instruction of NAS in more than one school talk about how much the organization differs between schools. Also, when welcoming students who move from other schools, they see differences in how grading has been done, often referring to "kindness grades." In CLIL, varying practices can be seen as well due to a lack of policy (Sylvén, 2019). The varying practices are a threat to equitable education for all, and are sometimes explained by the difficulty in implementing policy in the decentralized Swedish education system, although Sweden is perceived as a front-runner in bilingual education (Cummins, 2017).

Teacher education may be a good place to start. Ünsal (2017), referring to Cho and McDonnough (2009) and Lee (2005), mentioned in her thesis that one reason for science teachers' lack of knowledge about bilingual students' language development and learning in science is that the topic is rarely integrated into teacher education and curriculum development. The issue that needs to be addressed is what type of organization teachers are supposed to be molded into and prepared for. Below, some alternatives are outlined based on suggestions from teachers in Study II.

Suggestion 1: Content teachers as language teachers

One suggestion is to strengthen and develop the content teachers’ competencies in second language teaching methods. A couple of the NAS teachers comment that translators, study counselors and mother tongue instructors may be good, but what works best is teachers with a real understanding and familiarity with the subject:

- T6 [I]t is really good with those translations, [...] but to reach higher levels I think we rather need natural science teachers, because I think you get much further during the fifteen minutes I have after class with the students, even if I don’t know their language. I know what they need to know [...] and I think you achieve much better results that way, than seeing someone who knows their language but isn’t capable of explaining the natural sciences.
- T6 [B]ut sometimes you get an SSL teacher who knows the subject, and then it works so much better than having an SSL teacher who doesn’t know the subject, of course.

What the teacher above actually is advocating is for subject content teachers to be language teachers themselves, like the teacher below, who has a dual certification in SS and SSL.

- T10 As an SSL teacher you think about it automatically [...] you get used to reading texts from a second language perspective and bring up concepts automatically because you have a feeling they won’t understand this word, and they probably haven’t heard this concept.

This suggestion might involve a new kind of teacher education where all teachers are given at least some instruction in second language learning. In the CLIL context, one of the content teachers had taken university courses in English, but the rest simply found the ability to teach in English inspiring and saw it as professional development. In one case, it was not really a matter of choice, and the teaching in English required a lot of extra work but was perceived as rewarding in the end. Some CLIL teachers had received some input as regards teaching methods from EFL colleagues.

The language integration continuum, which was presented previously to define the teaching contexts, can be an instrument to decide how to organize instruction and define the relevant teacher profiles for both the CLIL and NAS contexts. The typical definition of CLIL and NAS in a Swedish setting could be positioned in the middle, with separate content and language classes, or as content-driven and thus positioned slightly to the left (see Figure 4.8). However, just because this is how they can be defined today does not mean that this is how it should be.

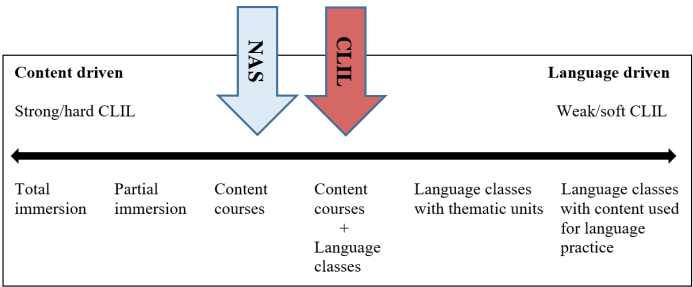


Figure 4.8 The position of Swedish CLIL and NAS contexts along the Language integration continuum

The corresponding EFL classes in CLIL did not typically integrate subject content but followed their own syllabus with course content to be covered. The teachers in Study II reported something similar as regards SSL, where colleagues could help with subject matter content in the beginning, but later on they had their own course content that had to be covered, leaving little time for subject integration.

Hence, the question is how teacher education responds to a need to equip subject matter teachers for language teaching. In December 2019, close to 80 professors, researchers and educators signed an article with the message that it is almost outrageous that language development is not yet a mandatory part of all teacher education programs. They referred to research saying that it is of utmost importance for multilingual students to have teachers who can teach language in all subjects, which is of extreme importance in today's segregated schools. An alternative solution or complement to teachers' dual competence is cross-disciplinary collaboration, which is outlined below.

Suggestion 2: Cross-disciplinary collaboration

Several of the teachers express a need to collaborate with colleagues, especially with Swedish language teachers, in the NAS context, but also subject content teachers within the same discipline. The teachers explain that they need help from SSL-teachers with ideas how to deal with language in their subjects, and disciplinary colleagues to work with co-assessment or in test development. In the CLIL context this need was hardly expressed at all. In one of the schools, with a stronger international profile the leadership had initiated that the teachers should study other subjects' syllabi to get acquainted with the content. The year it had been announced that they were going to start more cross-disciplinary collaboration.

Lack of time and sometimes lack of willingness among colleagues are perceived as obstacles that hinder collaboration on a regular basis in both contexts. Collaboration calls for a different organization to make it work, according to the teachers in both studies. The examples below come from Study II:

- R How can you as their teacher help them develop [both] the written and the oral language to [be able to] explain why [events happen]?
- T5 I believe we are not there yet in our team of colleagues. I think we **MUST** be more interdisciplinary.
- [...]
- T5 [T]he problem when doing these things is to find the means to do it, and time when we can collaborate, due to issues with the schedule, and the organization. That is part of the challenge as I see it if we are to have a really good language introductory program, we need an organization that can adjust to the short time frame that we must maximize.

The idea of collaborating with SSL teachers or mother tongue instructors was dismissed by some content teachers, who instead advocated the dual profile as seen above. They believed it is better with teachers who have a deeper understanding of the subject. Since the subject matter teachers are to grade the students, they need to know that the students have learnt the relevant content/language. Although collaboration between content teachers and language teachers represents an alternative path, this would require different organization at the schools and finding qualified SSL teachers, which is currently the teacher group with the lowest rate of certified teachers, according to statistics (National Agency for Education, 2018f). Collaboration thus requires rethinking the educational organization of teachers'

work. Other organizational changes were suggested as a result of both studies, however, and are presented below.

Suggestion 3: An alternative curriculum

When it comes to the relationship between time, content to be covered, language, and skills to be learned, one teacher in the introductory program notes that it takes a skilled student who might already have some study habits and is fast to pick up the language two years instead of one to finish a course in social studies. The teacher suggests there should be a “Swedish as a second language version” or a “light version” of the social science courses:

- T5 If there were a course plan with a second language-focus [...] it would have been so much better, because as it is now, they have to do everything [in a year or two] that other students do in nine years. The problem is that they are supposed to be assessed by the same measure as a Swedish elementary student, and it is so complex

The teacher comments that in a subject like social studies, it is a matter of achieving an understanding of foundational principles such as democracy. When asked what this means, the students answer “freedom, I can do as I want,” and the teacher has to explain that this is not correct; democracy means that the people decide, and you cannot always do as you want. Language is not the only issue, but conceptual understanding is closely tied to values and the understanding of meaning. Cultural perspectives should be included in a second language course plan in social studies, together with the very basics, according to the teacher.

The CLISS study in the CLIL context (Sylvén, 2019) generated suggestions in a similar vein, advocating a curriculum with specific language learning outcomes, which is already in place in similar language and content integrated programs in other countries (Coyle et al., 2009; Czura & Papaja, 2013). The goal is to provide equitable education and language by design rather than by default.

To conclude, the intention of this study was to describe and examine the policies and practices of language use in relation to the what and how of assessment and their possible consequences for fairness in access opportunities and assessment outcomes. In this study I take a perspective that equitable assessment requires that teachers assess students in ways consistent with how they were taught (cf. p.28) and that validity in assessment is inseparable from fairness in access. There are still questions that need to be resolved in relation to both policy and practices, e.g. language use, organization and collaboration at the schools and comparability and validity in the use of different assessment formats. More research in the field of assessment is clearly needed, and some suggestions are provided below.

Suggestions for future research

Hopefully the future will bring a broad body of research, helping build equitable assessment practices. Starting with the participants, I welcome research from a student perspective, using interviews, observations and statistical analyses of assessment results, as well as experimental designs investigating experiences or effects of different assessment methods and accommodations. In agreement with what others have already proposed (Abell & Siegel, 2011), I suggest more research be conducted on the relationship between assessment practices in the classroom and student learning to see whether teachers with more visible language pedagogy and aligned assessment practices help students perform better. Think-aloud protocols could be used while students do different assessment tasks.

In relation to CLIL and English-medium instruction, it would be interesting to compare student performance during their early schooling, when first exposed to instruction in English, with performance in Grade 9, before going to upper secondary. It would also be interesting to look at case studies and individual student trajectories to compare how students perform in Swedish versus English at upper secondary school or in college after doing the NS in English and the SS in Swedish, as is the case in some international compulsory schools. What are the students' experiences concerning the different languages of instruction in the subject disciplines?

Below is a list of suggestions, organized thematically; collecting bigger data (more of the same), additional data (from other contexts to enable comparison) and with the use of other methods.

Bigger and richer data (including more of the same):

- Teachers in schools with extensive experience with second-language learners and migration (e.g., in suburban areas which was scarce in this study).
- More teachers in the social and natural sciences at each level, in order to see if results correlate with those in this study (e.g. if the practices of teachers with a dual language/content subject combination alters teachers' assessment beliefs and practices).
- Investigating oral assessment practices in the social and natural sciences.

Additional data (reaching similar yet different contexts to enable comparison):

- Teachers in other subject matter domains (e.g., practical and vocational subjects).
- Private international English-medium instruction schools.
- Tertiary education.
- More language teachers, SSL in particular, to be able to investigate collaboration and possible interdisciplinary assessment content and methods.
- Look at assessment practices in other countries with, for instance, similar migration situations.

Different methods (as in multilevel research, taking different micro-, meso- and macro-levels into account):

- Investigating and comparing different groups' assessment beliefs: students, teachers, student counselors, parents/guardians, policy makers, politicians, student teachers, teacher educators.

Different methods (as in adding or shifting focus):

- Classroom observations to compare teaching methods and assessment practices.
- Longitudinal studies which explore how teachers' conceptions of assessment influence their assessment literacy development. In particular, evidence if and in that case how assessment training can change teacher conceptions toward a more learning-facilitating direction would be welcome, according to Xu and Brown (Xu & Brown, 2016).
- Experimental studies examining the effects of different assessment methods and accommodation strategies.
- Document analyses analyzing/comparing policy documents and assessment tasks/tests.

In order to safeguard equitable, valid and fair assessment practices, more research is needed across all the fields indicated above to inform teacher education and language policy. Assessment in a heterogeneous multilingual context requires empirically based knowledge, and a lot is to be gained in communication and exchange with teachers and students, who possess a great deal of important experiences and expertise.

Summary in Swedish/Svensk sammanfattning

Bedömning i flerspråkiga skolor: En jämförande mixad metod-studie av lärares syn på och arbete med bedömning bland språkinlärare – CLIL och nyanlända elever

Inledning och bakgrund

Rättvis, rättvisande och likvärdig bedömning är det samma som rättvis tillgång till ämnesstoff, undervisning och lärande. I en flerspråkig undervisningsmiljö, där elever förväntas tillägna sig ämnesstoff samtidigt som de lär sig undervisningsspråket blir språkets roll central för validiteten i bedömningen. Lärare behöver vara medvetna om de möjliga effekter som elevers otillräckliga språk kan få i olika bedömningsuppgifter och aktiviteter., För att ge elever lika förutsättningar är det viktigt att synliggöra nödvändiga språkliga register i undervisningen samt att överväga olika typer av anpassningar och stöttning i samband med exempelvis prov (Abedi, 2008, 2009; Llinares et al., 2012; Sireci & Faulkner-Bond, 2015; J. W. Young, 2009). Likvärdig bedömning är därmed beroende av hur undervisning och bedömning är uppbyggda samt samstämmigheten dem emellan (Kane, 2010; Stobart, 2005). Det handlar om att elever ges tillfälle att utveckla inte bara sin förståelse utan även de ämnesspecifika språkliga uttrycksformerna såväl som det allmänna akademiska skolspråket (Cummins, 1979; Schleppegrell, 2001, 2004, 2011). I svensk skollag 1 kap. 4 § (SFS 2010:800) talas det om skolans kompensatoriska uppdrag, att ge alla elever möjlighet att nå så långt som möjligt och kompensera för elevers olika förutsättningar, ett uppdrag som är en utmaning i en alltmer språkligt heterogen skola.

I svensk skola är svenska som undervisningsspråk norm. Samtidigt finns internationella grundskolor där upp till hälften av undervisningen bedrivs på ett annat språk, företrädesvis på engelska. Omfattningen är reglerad i Skolförordningen (SFS 2011:185, 9 kap. 18-19§§) som säger att maximalt upp till hälften av undervisningen kan bedrivas på engelska. Internationellt benämns dessa skolor ofta som CLIL, *content and language integrated learning*. På gymnasienivå, där engelskspråkig undervisning blivit vanligare som ett sätt att locka elever (Hennerdal et al., 2018; National Agency for Education, 2018h; Sylvé, 2019), är språkanvändningen oreglerad vilket är unikt för Sverige och har ifrågasatts både i forskning (Sylvén, 2019) och i en rapport från Skolverket (2018h). Inom högskoleutbildning har engelskans roll som undervisningsspråk fortsatt att öka i betydelse, inte minst i samhällsvetenskapliga discipliner (Dalberg, 2013). Detta har en direkt effekt som kan diskuteras även i denna studie som är skriven på engelska vilket innebär att Skolverket i referenser benämns som *National Agency for Education*). Inom naturvetenskapliga ämnen har engelskans ställning varit stark sedan länge och betraktas som ett statusspråk där över 90% av uppsatserna på avancerad nivå skrivs på engelska (Salö, 2010). Detta kan ses som ett argument för den ökade förekomsten av engelskspråkiga program i gymnasieskolan.

Sverige har en lång erfarenhet av migration, men mellan 2012 och 2016 fördubblades antalet nyanlända i grundskolan (Mörthund, 2020). Enligt statistik från Skolverket är numera var fjärde elev i grundskolan av utländsk bakgrund och nära var tredje i gymnasieskolan (National Agency for Education, 2019b, 2019c). Läsåret 2017/18 visade Skolverkets statistik att det förberedande programmet språkintrödn

var det största enskilda programmet på gymnasiet i antal studenter. Redan 2014 noterade OECD (2014) att Sverige är ett av de länder som har störst andel utlandsfödda och att integration därmed är avgörande för både svensk ekonomi och för samhället i stort. En statlig utredning tillsattes för att utreda hur nyanländas möjligheter till skolframgång kan förbättras (SOU 2017:54), detta efter rapporter om att endast 16% av sent anlända elever som kommit i grundskolans årskurs 6-9 blev behöriga att söka nationella program på gymnasiet 2016, två år senare var andelen 28% (National Agency for Education, 2018f).

I Skolverkets råd för undervisning av flerspråkiga elevgrupper och nyanlända elever rekommenderas användning av elevens starkaste språk initialt liksom möjlighet till transspråkande, att blanda språk (e.g. National Agency for Education, 2012b, 2016, 2018e, 2018i). Fördelen med transspråkande och tillgång till elevers första språk för nyanlända är något som påtalats i forskning inom området (García & Wei, 2018), inte minst för att stärka elevernas identitet. Berglund (2017) noterar att Sverige skiljer sig från många andra länder och väljer integration snarare än assimilation och ser tvåspråkighet som en resurs. Sverige har även utpekats som ett föregångsland på området med en väl utbyggd modersmålsundervisning vilket sträcker sig långt tillbaka i tiden (Cummins, 2017). Nyanlända elever har rätt till studievägledning på sitt modersmål enligt Skollagen (SFS 2010:800, 5 kap. 4§). Däremot har det visat sig att i svensk skola som präglas av decentralisering finns stora variationer mellan kommuner och skolor i hur detta genomförs, vilket även gäller den lagstadgade kartläggningen av nyanländas litteractitet (SOU 2017:54). Kartläggning är obligatorisk att genomföra under de första två månaderna enligt skollagen (SFS 2010:800, 3 kap. 12 §) i syfte att fatta beslut om placering av elever och för att upprätta en individuell studieplan (National Agency for Education, 2016) men det har visat sig att hur och om den genomförs skiftar betydligt.

Brister i likvärdig betygsättning och bedömning mellan skolor och lärare har uppmärksammats som ett problem i svensk skola inte minst beroende på ökad segregation och en växande klyfta mellan gynnade och missgynnade elever. OECD (2019a) pekar på att klyftan är större i Sverige än OECD-genomsnittet. Ökande migration och effekter av skolvalsreformen anges som bidragande orsaker. OECD (2017) pekar även på behovet av att stärka lärarprofessionen då det har varit svårt att behålla lärare i yrket, särskilt behöriga. Efter 2015 har många små kommuner som tidigare haft få eller inga elever med svenska som andraspråk tagit emot många nyanlända vilket ställer stora krav på lärare och skolans organisation. Mindre kommuner har samtidigt svårare att rekrytera behöriga lärare. Kartläggningen av engelskspråkiga gymnasieskolor som Skolverket fick i uppdrag av regeringen att utföra visar att andelen lärare med behörighet att undervisa i olika ämnen är lägre där engelska är undervisningsspråk än för genomsnittet. I biologi exempelvis är 84% av lärarna på samtliga gymnasier behöriga jämfört med 64% på engelskspråkiga program (National Agency for Education, 2018h). Sett till hela lärarkåren är behörigheten lägst bland lärare i svenska som andraspråk. (National Agency for Education, 2019d).

Medan svenska elever ligger i topp i EF Proficiency Index (EPI), en ranking baserat på engelsk språknivå i länder och regioner som inte har engelska som modersmål, så är engelska ett ämne som nyanlända generellt har svårast för, förutom svenska (SOU 2017:54). Medan CLIL-program samlar främst studiemotiverade elever som i många fall ser engelska som en dörr utåt för internationella möjligheter så är målet för nyanlända att hitta dörrar in i svenska samhället. Medan CLIL-elever oftast undervisas av lärare som själva har engelska som andraspråk, undervisas nyanlända oftast av svenska modersmålstalare. För nyanlända är svenska målspråk, men för CLIL är målspråket helt eller delvis engelska. CLIL-elever har oftast en god eller mycket god språklig kompetens redan från början, men nyanlända är nybörjare i undervisningsspråket. CLIL-elever får sin språkliga exponering genom så kallade extramurala fritidsaktiviteter, nyanlända får i bästa fall sin från omgivningen, om de inte bor i

en språkligt segregerad miljö. Oavsett kan inte elever, CLIL eller nyanlända, förväntas få akademisk språklig input i målspråket utanför skolmiljön. I CLIL-kontext konstaterar lärare att även om eleverna är språkligt kompetenta så gäller det inte de akademiska registren i samma utsträckning vilket kommer påverka betygen negativt (Reierstam, 2015).

Mot den bakgrund som beskrivits ovan, och med utgångspunkten att tillgång till språk är avgörande för likvärdiga utbildningsmöjligheter och rättvisande bedömning, samt att forskning kring bedömning i flerspråkig miljö är begränsad, växte denna avhandling fram. Den består av två delstudier. Studie I genomfördes 2012-2015 inom ramen för forskarskolan FRAM för yrkesverksamma lärare och utmynnande i en licentiatuppsats som undersökte lärares syn på bedömning och bedömningspraktik på engelskspråkiga gymnasieskolor, så kallade CLIL-skolor (Reierstam, 2015). I den första studien ingick lärare i biologi, historia och engelska för att kunna jämföra lärare som använde svenska respektive engelska som undervisningsspråk samt kollegor som undervisade engelska på samma skolor. Detta för att se om användningen av engelska i ämneskurser påverkade lärares sätt att tänka kring och arbeta med bedömning och språk både inom kurser och mellan kurser. I Studie II som genomfördes 2017-2020 var fokus att beskriva ämneslärares syn på bedömning bland nyanlända (NAS, newly-arrived students) och i den inledande storskaliga enkäten ingick ett bredare spektrum av lärare än i den första studien. Lärare i flera samhälls- och naturvetenskapliga ämnen inbjöds att delta från tre skolformer: årskurserna 7-9 på grundskolan, introduktionsprogrammet och gymnasieskolan, detta för att möjliggöra en bredare datainsamling (se nedan under metod och material). Språklärarperspektivet som fanns med genom engelsklärare i Studie I, avgränsades här till vissa lärares dubbla behörighet, ämne och ett språk.

Syfte och forskningsfrågor

Det övergripande syftet med den här avhandlingen är att bidra till forskningen kring bedömning i flerspråkig miljö genom att undersöka lärares syn på bedömning bland språkinlärare. Lärares tysta kunskap beskrivs ibland som en svart låda men kan bli synlig genom deras självreflektion och i deras självkonstruerade bedömningsuppgifter. Avsikten är att beskriva språkpolicy och språkanvändning i relation till bedömning och på så vis dra uppmärksamhet till hur olika policy och varierande förhållningssätt till språk i undervisningen kan påverka validitet, rättvisa och jämförbarhet i bedömning av ämneskunnande bland språkinlärare. Följande forskningsfrågor användes:

Övergripande forskningsfrågor Studie I och II:

- Vad har lärare för syn på bedömning och bedömningspraktik bland språkinlärare? Skiljer de sig åt beroende på kontext, CLIL jämfört med nyanlända? I så fall, hur skiljer de sig åt?
- Vad har lärare för syn på språk i relation till ämnet och policy? Finns det gemensamma drag mellan CLIL och nyanlända-kontext? I så fall, vilka är dessa?

Forskningsfrågor Studie I:

- Finns det skillnader i bedömningspraktik beroende på om undervisningsspråket är engelska eller svenska, CLIL respektive icke-CLIL, i biologi och historia på gymnasiet? I så fall, *hur* skiljer den sig i så fall åt och *på vilka grunder*?

- Påverkas innehåll och bedömningsformer i engelskkursen i de fall där andra ämnen undervisas på engelska? I så fall, *hur* påverkas de?
- Hur ser bedömningsformerna ut i de olika disciplinerna med avseende på *språk, form* och *innehåll*? Finns det gemensamma drag?

Forskningsfrågor Studie II:

- Vad är ämneslärares självrapporterade syn på bedömning och bedömningspraktik bland nyanlända elever?
- Hur hänger lärares självrapporterade bedömningspraktik samman med deras syn på eleverna och språkpolicy i utbildning, (närmare bestämt i läroplan, ämnes-/kursplaner, transspråkande praktiker och lokal organisation)?
- Påverkas lärares syn på bedömning och bedömningspraktik beroende på skolform och ämne? Om så är fallet, hur påverkas de och på vilka grunder?

Forskningsbakgrund och teoriram

Forskning om bedömning i flerspråkig miljö är fortfarande tämligen begränsad, i CLIL-sammanhang beskrivs den till och med som ”en vit fläck”. Motivet är att hitta en pedagogik och bedömningsformer som gör att elever inte missgynnas vid bedömning av ämneskunnande på grund av språket och att i förväg definiera språkliga lärandemål (Hönig, 2009; Kiely, 2009; Llinares et al., 2012; Massler et al., 2014) för att främja validitet och rättvisa. Messick (1989) beskriver svårigheten att kunna dra valida slutsatser av provresultat i de fall språkliga kompetenser som eleverna inte har fått tillfälle att tillägna sig påverkar bedömningens utfall. Detta benämner han *konstrukt irrelevant varians* vilket uppstår om ett prov är utformat så att det exempelvis bedömer läsförståelseförmåga istället för ämneskunnande. Frågor rörande otydlig språkpolicy och bristande överensstämmelse mellan det språk som används i undervisningen respektive vid bedömning kommer i fokus, vilket i CLIL-sammanhang berörs av bland annat Gablasova (2014). Det är även ett fokusområde i forskning om bedömning bland andraspråksinlärare i majoritetsspråket och att göra kunskapsstoff tillgängligt för elever med annat modersmål av rättviseskäl (Abedi, 2009; Kane, 2010; Schleppegrell, 2004, 2005, 2011; Sireci & Faulkner-Bond, 2015; Stobart, 2005; J. W. Young, 2009).

Eftersom de två delstudierna i den här avhandlingen är flerdimensionella och inbegriper lärares uppfattningar, bedömningsteori, de olika ämnesdisciplinernas karaktärsdrag, en funktionell språksyn och språkpolicy uttryckt i bland annat kurs- och ämnesplaner, blir studiens teoretiska ramverk ganska brett. Nikula, Dalton-Puffer och Llinares & Lorenzo (2016) har använt en modell för att illustrera den sammansatta komplexiteten i CLIL där det integrerade förhållandet mellan språk och innehåll får implikationer på flera nivåer. Modellen som har anpassats för och använts som illustration även i den här studien har tre komponenter: *läroplan och policy*, *läraren* och *bedömningspraktiken*. De tre komponenterna kan även översättas med, *vad*, *vem* och *hur*. Läraren omsätter kursmål och rådande policy i praktik, i detta fall mer specifikt i bedömningspraktik även om bedömning här ses som en integrerad del av undervisningspraktiken.

Relationen mellan förväntade lärandemål och undervisningens upplägg finns beskriven i Biggs (2003) modell för konstruktiv länkning, vilket i sin tur för vidare till ytterligare en annan modell som använts som referens i syntesen av den här studiens resultat och kan översättas med "läroplanslänkningsmodellen". Tre nivåer hänger samman och utgör undervisningens helhet: *planerad läroplan*, *implementerad läroplan* och *uppnådd läroplan* (Porter, 2002; Thijs & Akker, 2009). Den planerade läroplanen är oftast synonymt med den nationella läroplan som är fastställd av regeringen medan den implementerade är det som omsätts av läraren i planering och undervisning. Den uppnådda läroplanen representeras sedan av vad elever har lärt sig, vilket inte nödvändigtvis överensstämmer med de två föregående. I diskussionen av de två delstudierna, CLIL och NAS, jämförs vilka riktlinjer som finns avseende språkanvändning i policydokument och hur detta påverkar de deltagande lärarnas praktik. Utifrån detta förs ett resonemang kring vad det i sin tur kan få för konsekvenser för statusen hos olika språk, för lärandet i de fall språket förväntas komma av sig självt och vad som de facto anses vara målspråk i de två kontexterna, svenska och/eller engelska och/eller modersmål?

En annan modell som var användbar för syntesen av studiens resultat förenar graden av synlig språkpedagogik med graden av målstyrning. Modellen är utvecklad av Leung och Morton (2016) i samband med deras syntes av forskningsstudier inom CLIL. Leung och Morton skapar en fyrfältsmatris genom att horisontellt använda sig av Bernsteins (2000) koncept synlig respektive osynlig pedagogik, där det första karaktäriseras av uttrycklig språkundervisning och det sistnämnda av naturlig språkinläring. Vertikalt använder Leung och Morton sig av Bakhtins (1981) idé om spänningen mellan större eller mindre central styrning, alternativt uttryckt som normstyrt eller diversifiering. I det här fallet handlar det om rådande språknorm och hur styrande kursmål och synen på ett gemensamt målspråk är. En undervisning med stark betoning på att undervisa inte bara koncept utan även annat akademiskt skolspråk och som utgår från att det finns ett tydligt definierat målspråk uttryckt i kursmålen hamnar högt upp till höger i modellen. En undervisning som har stark betoning på språkutveckling men i dialog, utan att föreskriva ett speciellt språkbruk, hamnar långt ner till höger i modellen och omfattar exempelvis transspråkande praktiker där diversifiering är ett kännetecken. Medan det första exemplet utgår från större språklig styrning, utgår det andra från lägre grad av styrning (cf s. 41f).

Gemensamt för Studie I och II var fokus på funktionsorden i lärandemålen (cf Bloom, 1956; Dalton-Puffer, 2016) och deras förväntade språkliga och kognitiva krav på elevers prestation i bedömningsuppgifter. Funktionsorden graderas ofta med betäckningen "lower order" respektive "higher order thinking skills" (Anderson & Krathwohl, 2001; Verma, 2014) och kategoriseras därmed längs ett kontinuum där mindre krävande kan exemplifieras med *att återge* medan *att analysera* representerar mer krävande kognitivt processande. Att resonera och analysera kräver i allmänhet även mer språklig förmåga och språk av en viss karaktär. Funktionsorden finns med Cummins matris (1981), och Hall's (1995) vidareutveckling av modellen, som användes för att beskriva och analysera de språkliga kraven i lärarnas bedömnings exempel. Ett flertal andra verktyg användes i analyserna av relationerna mellan språk och innehåll samt lärarnas bedömningsfokus: *varför*, *vad*, *hur* bedömer lärare och vad får det för konsekvenser, "och?" Vad vilar det på för syn på eleverna, syn på lärande, värderingar och principer, utifrån en "bedömningslitteracitetsmodell" utvecklad av Abell och Siegel (2011). Utgångspunkten för synen på bedömning i den här studien är som redan nämnts att bedömning är en integrerad del av undervisningen. Det innebär också en uppfattning att den summativa och formativa bedömningen måste fungera tillsammans, även om fokus i den här studien är framför allt på summativ bedömning. Därmed blev valideringsfrågor centrala avseende rättvis förberedelse och stöttning samt relevans i relationen mellan målet för bedömningen, konstruktet, och bedömningsuppgifternas format (Crooks et al., 1996; Kane, 2010; Messick, 1989; Sireci & Faulkner-Bond, 2015).

All data i den här studien grundar sig på lärares utsagor och exempel på deras egenkonstruerade bedömningsuppgifter, som redan nämnts. Det innebär att den teoretiska ramen vilar på "teacher cognition", lärarkognition (Studie I) och "teacher beliefs", lärares uppfattningar (Studie II). Teacher cognition utgör en gren av teacher beliefs och är framför allt vanligt vid beskrivning av språklärares praktik (Borg, 2003). I forskning har betydelsen av lärares uppfattningar (beliefs), alternativt uttryckt som pedagogiska grundsyn, noterats påverka lärares undervisningspraktik (Fives & Gill, 2014) och förväntas grunda sig på lärares utbildning och samlade erfarenhet av undervisning (J.-L. Berger et al., 2018; OECD, 2009a). Inom bedömning används termerna beliefs, conceptions och perceptions växelvis (N. Barnes et al., 2015; G. T. L. Brown, 2004, 2006; Harris & Brown, 2009; Remesal, 2007). Lärares "beliefs" inom tre områden anses vara särskilt viktiga för lärarpraktiken, nämligen deras syn på undervisning, deras kunskapssyn och slutligen deras syn på elevers förmåga (Fives & Buehl, 2016). Dock finns studier som pekar på inkonsekvenser mellan lärares pedagogiska synsätt och praktik (Pajares, 1992) vilket ofta anses vara ett resultat av påverkan av omgivande faktorer (Li, 2013; Poulson et al., 2001; Tamimy, 2015).

Metod och material

Materialet för den första studien som insamlades under 2013-14, består av halvstrukturerade intervjuer, en enkät samt bedömningsmaterial i form av skriftliga provexempel. Till datainsamlingen i den andra studien under 2018-2019 användes en storskalig webbenkät, uppföljande halvstrukturerade intervjuer och precis som i Studie I lärares prov och bedömningsuppgifter. Intervjuerna transkriberades och analyserades tematiskt vilket kan benämnas som "kodbok-tematisk analys" (Braun et al., 2019) där vissa teman är givna i förväg baserat på intervjuguiden exempelvis och kursmål medan andra framträder vid analysen. De skriftliga prov som de intervjuade lärarna delade med sig av analyserades utifrån faktorer såsom frågetyp, t ex flervalsfrågor, matchningsfrågor, kortvarsfrågor eller uppsatsfrågor (J. D. Brown & Hudson, 2002; Wikström, 2014) samt vilka krav proven ställer på elevers förväntade kognitiva och språkliga förmåga. Blooms' reviderade taxonomi (Anderson & Krathwohl, 2001) och så kallade akademiska funktionsord i provfrågor var en del av analysverktygen (se ovan). Likaså användes en tillämpning av Cummins' kvadrant som kombinerar kognitiv komplexitet med språklig komplexitet. I Studie II användes programvaran SPSS Statistics för de statistiska analyserna vilka inkluderade både deskriptiva och inferentiella analyser. Enkäten bestod av nominal- och ordinaldata varför icke-parametriska tester gjordes för att analysera effekter på gruppnivå. Mann-Whitney användes för signifikansprövning mellan två grupper (So och No-lärare) och Kruskal Wallis användes för tre grupper (tre olika skolorformer).

Medan den första studien är huvudsakligen kvalitativ, användes mixad metod i Studie II där kvantitativ och kvalitativ analys användes i en *interaktiv mixad design* där komponenterna kontinuerligt jämfördes med varandra och utvecklades under processen (Schoonenboom & Johnson, 2017). De fem komponenterna utgörs av syfte, konceptuellt ramverk, forskningsfrågor, metod och validitet (Maxwell & Loomis, 2003). Studien befinner sig inom fältet utbildningsvetenskap och pedagogik och representerar praxisnära forskning. Fältet präglas ofta av ett pragmatiskt förhållningssätt där fokus är på hur forskningsfrågorna bäst kan besvaras. I detta fall innebär det en kombination av ett deduktivt och induktivt förhållningssätt, även det interaktivt, där jag som forskare rörde mig fram och tillbaka mellan observation, tolkning och teoretisk förståelse (Bryman, 2016). De teoretiska utgångspunkterna i en studie som denna är flerdimensionella som konstaterats ovan men utgångspunkten är en kombinerad essentialistisk och konstruktivistisk språksyn, där jag som forskare ser språk som system med givna

komponenter men som sätts samman och konstrueras i användning. Detta kan beskrivas som en systemisk funktionell språksyn (Halliday, 2014) där varje språk är ett meningsskapande system som växlar i sina uttryck beroende på användning i olika kontexter (Schleppegrell, 2020). Denna kombinerade språksyn överensstämmer med kunskapssynen i bedömningssammanhang där bedömning sker mot föruppställda mål med förväntade ämnesgenrer, fast konstruerat av eleven.

Studie I

Studie I omfattar tolv lärare, varav tre undervisar i biologi, tre i historia och resten i engelska vid tre gymnasieskolor i en stor samt två mellanstora svenska städer. En av skolorna är helt engelskspråkig, men med svensk läroplan, de andra två erbjuder ett respektive två gymnasieprogram med internationell profil där undervisningsspråket i övervägande fall är engelska. Denna studie ingick i ett större projekt, finansierat av vetenskapsrådet, CLISS-projektet, Content and Language Integration in Swedish Schools. För en närmare beskrivning, se Sylvé (2019) och Sylvé & Ohlander (2014). De aktuella skolorna och lärarna ingick redan i den aktuella forskningsstudien och kontakter behövde inte knytas.

En intervjuguide användes med följande teman: lärares erfarenheter av CLIL, ämnessyn, syn på bedömning samt använda bedömningsinstrument, kursmaterial, kursplan och förekomsten av ämnesövergripande samarbete. Enkäten som följde efter intervjuerna syftade till att ge större jämförbarhet mellan svar då den fria semistrukturerade intervjuformen gjort att vissa frågor visade sig svåra att jämföra på grund av undvikande svar.

Studie II

I Studie II genomfördes en storskalig enkät (N=196) bland lärare i kommuner som identifierats med omfattande mottagande av nyanlända elever efter 2015. Fokus var skolor med stort antal sent anlända elever från årskurs 7 och uppåt eftersom utmaningarna för dessa elevgrupper att tillägna sig språk och ämne samtidigt är som störst. Eftersom det visade sig svårt att få deltagare, delvis eftersom kontakten med lärare gick via rektorer, utökades urvalsgruppen som grundade sig på bekvämlighetsurval, allt eftersom. Det initiala syftet var att hitta grupper av lärare med olika lång erfarenhet av nyanlända språkinlärare. Tanken var att jämföra lärare i storstadsområden respektive i mindre kommuner med en nylig inflyttning av stora grupper av nyanlända. Det visade sig dock svårt att rekrytera lärare i storstadsregionerna i synnerhet. Flera rektorer angav att lärarna redan är med i andra studier. Kommunerna som kontaktades valdes ut baserat på migrationsverkets statistik över kommuner med stort mottagande av nyanlända under åren 2015-2017. Det slutliga urvalet landade på 20 deltagare från storstadskommuner, 78 från mellanstora kommuner eller pendlingskommuner och 89 från mindre städer och landsbygdskommuner (n=187). Några ville inte ange sin jobbkommun (n=9). SKRs definition av kommuntyper från 2017 användes. Jämförelser med nationell statistik visade att urvalet någorlunda motsvarar det nationella genomsnittet men medel för undervisningserfarenhet var något högre för deltagarna i enkäten, nästan 15 år jämfört med snittet för riket på mellan 12 och 13 år. Likaså låg snittet för lärare med behörighet något högre, 93% av deltagarna jämfört med 85-90% för riket beroende på ämnestillhörighet.

I enkäten kunde deltagarna ange om de var intresserade av att delta i en uppföljande intervju, ett 20 tal lämnade sina kontaktuppgifter, men efter ett urval baserat på aktuell undervisningserfarenhet i samhälls- och naturvetenskapliga ämnen och visst bortfall slutade det med 13 lärare som deltog i intervjuer. Till

intervjuerna ombads lärarna ta med sig några bedömningsexempel att diskutera utifrån deras erfarenheter med nyanlända i någon kurs och alla lärare tog med sig några exempel. Av de deltagande lärarna representerade 10 So-ämnen och 3 No-ämnen. Lärarna kom från olika skolformer, 6 som vid tidpunkten undervisade i grundskolans årskurs 7-9, 5 på introduktionsprogrammet varav 3 även var på gymnasiet och 2 som endast var på gymnasiet. En intervjuguide användes i likhet med Studie I som inkluderade frågor om lärarnas bakgrund, elevgruppernas sammansättning, eventuellt samarbete med kollegor, lokal språkpolicy, språk i undervisning och bedömning samt bedömningsexemplens utformning.

Resultat

Studie I

I förhållande till den första forskningsfrågan: *Finns det skillnader i bedömningspraktik beroende på om undervisningsspråket är engelska eller svenska, CLIL respektive icke-CLIL, i biologi och historia på gymnasiet? I så fall, hur skiljer den sig i så fall åt och på vilka grunder?* föreföll det inte som att CLIL och användandet av engelska som undervisningsspråk hade någon särskild inverkan på lärares bedömningspraktik. De skillnader som kunde spåras, framför allt i historia, verkade snarare bero på individuella preferenser eller ämnets karaktär. De två CLIL-lärarna föredrog uppsatser, men medan den ena förordade en narrativ genre, då detta bäst ansågs avspegla den historiska ämneskaraktären, föredrog den andra en mer akademisk form, vilken ansågs vara mer tydlig och strukturerad. En kollega (icke-CLIL) föredrog frågeprov, men vissa av frågorna, framför allt mot slutet av kursen, hade uppsatskaraktär.

I biologi användes både frågeprov och andra skrivuppgifter såsom laborationsrapporter och fältanteckningar. Två kollegor, en CLIL och en icke-CLIL vid samma skola, samarbetade kring utformningen och flera provfrågor och uppgifter var närmast identiska. En skillnad som framkom i intervjuerna, var att CLIL-läraren fokuserade ämnesspecifik vokabulär, vilket delvis kan härledas till användandet av engelska, men även icke-CLIL läraren försåg eleverna med ordlistor över ämnesspecifika begrepp. CLIL-läraren medgav att det generella akademiska utfyllnadsspråket inte ägnats någon särskild fokus, men påpekade dock att det skulle vara en intressant aspekt att ha i åtanke. Icke-CLIL kollegan däremot nämnde behovet att tillgodose utvecklandet av ett mer generellt akademiskt språk. Sammanfattningsvis kunde vissa modifieringar i bedömningspraktiken på grund av CLIL spåras; frekvensen av frågeord på en lägre kognitiv nivå kunde förefalla något högre i CLIL-prov och eleverna hade i vissa fall möjlighet att välja vilket språk de ville använda, ha tillgång till ordböcker eller få längre tid på sig vid prov. Dessutom fanns möjligheten att göra förtydliganden i efterhand om språket tycktes utgöra ett hinder. Alla CLIL-lärare gav uttryck för att det enda som bör skilja sig åt är språket, och hänvisade till att de ju lyder under samma nationella ämnesplaner och kursmål. I enkäten angav alla lärare utom en, att nationella kursmål är mycket viktiga vid betygssättning.

Svaret på den andra forskningsfrågan, om innehåll och bedömningsformer i engelskkurserna påverkas, var i huvudsak nej. Frågan som infinner sig är vad bristen på avspeglning av innehåll från ämneskurserna beror på? Enligt lärarna själva berodde det på brist på tid att hinna med något annat än det kursinnehåll de är tvungna att klara av i relation till sina ämnesmål. Engelsklärarna fäste stor vikt vid de nationella proven. Dessa användes som referens för vad som behöver ingå i kursen, och gamla prov användes som bedömningsinstrument och som förberedelse för de "riktiga" proven. En annan faktor som hämmade

ämnest integrering beskrevs som brist på tid för samarbete och samplanering. Slutsatsen är att CLIL verkar hålla sig inom ämneskurserna, vilket bekräftas av annan forskning (Massler et al., 2014).

Den sista forskningsfrågan berör gemensamma nämnare mellan ämnena avseende språk, innehåll och form i bedömnings sammanhang. Den främsta likheten fanns i användandet och betoningen av skriftliga bedömningsformer vilka medför ett behov av att eleverna behärskar både ett ämnesspecifikt och ett allmänt akademiskt språk. En skillnad som framkom berörde lärarnas uppfattning om elevernas språkliga kompetens. Medan engelsklärarna såg CLIL-eleverna som avancerade språkinlärare, ansåg ämneslärarna att de måste stötta eleverna språkligt. Detta kan sägas återspegla att två olika språkliga register är i fokus i de två kontexterna, vilket kan jämföras med Cummins BICS, Basic Interpersonal Communicative Skills och CALP, Cognitive Academic Language Proficiency (Cummins, 2008), där det första är mer frekvent i engelskklassrummet, och det sistnämnda i ämneskurserna. Analyserna visade att det fanns vissa skillnader mellan disciplinerna avseende vad som kan beskrivas som innehåll. En beröringspunkt relaterade till det gemensamma akademiska språket och användandet av skriftliga genrer. Det ämnesspecifika innehållet skulle dock kunnat användas för att berika det språkliga uttrycket i engelskkurserna. Begreppet innehåll visade sig vara kopplat till ämnets karaktär. I biologi brukar ämnesbegreppen beskrivas som en del av ämnesinnehållet (Linares et al., 2012), men detta är inte lika tydligt i historia.

Studie II

En sammanfattande slutsats i enlighet med den första forskningsfrågan om ämneslärares självrapporterade syn på bedömning och bedömningspraktik bland nyanlända elever, är att en majoritet av lärarna ansåg att bedömning är svårt och de uttryckte att utmaningen vid bedömning är språket eftersom nyanlända elever har begränsad förmåga att uttrycka sitt ämneskunnande. Liksom i Studie I varierade lärares bedömningsformer beroende på ämnets karaktär, men alla refererade till kunskapskraven i sina respektive ämnen på frågan om vad som bedöms. Här är funktionsorden viktiga och ämnesspecifika förmågor, att kunna resonera och dra slutsatser exempelvis, vilket angavs som det viktigaste vid bedömning i ämnena. I undervisningen beskrev lärarna däremot att fokus ligger framför allt på ämnesspecifika begrepp, men vissa lärare angav att de jobbade med modelltexter och att lära ut användbara formuleringar. Några av de intervjuade lärarna beskrev att det sistnämnda har kommit mer i fokus på sista tiden och att de har det som en gemensam strategi antingen i ämnesgruppen eller på skolan. I enkäten rapporterade lärarna att eleverna har särskilt svårt för ämnesspecifika genrer, såsom laborationsrapporter eller utredande uppsatser. Det gäller skriftlig såväl som i muntlig form, men särskilt skriftlig. Medan lärarna, ibland något ovilligt, menade att skriftliga bedömningsuppgifter var i fokus, bl. a. med motiveringen att elever behöver lära sig skriva för att klara av högre studier, sågs muntliga bedömningsformer som en möjlig anpassning som även kunde användas som uppföljning då elever inte klarat det skriftliga provet.

I relation till den andra forskningsfrågan och lärarnas syn på eleverna, beskrev flera lärare att det märks att de nyanlända eleverna saknar exponering för vardagsuttryck medan de däremot kan många ämnesspecifika ord som de mött i undervisningen. Elevernas bristande referenser till vardagliga fenomen angavs också som ett bekymmer vid bedömning då elever inte har förståelse för vad det innebär att exempelvis ”gunga gungbräda”. Likaså sade lärarna att det är svårt att förklara för eleverna att det inte räcker att memorera och återge fakta vilket NAS verkar vana vid från sin tidigare skolgång. Eleverna har svårt att först och främst ta till sig vad kritiskt tänkande innebär, som en lärare beskrev det: ”Va, får jag tycka något annat än vad boken eller läraren säger?”, sedan tillkommer svårigheten att ha tillräckligt

med språk för att uttrycka kritiskt tänkande. Användningen av resultat från de lagstadgade kartläggningsskärmen för nyanlända för att få förståelse för elevernas literacitet var tämligen låg, vilket överensstämmer med rapporter från bland annat den statliga utredningen (SOU 2017:54).

Lärarna återkom till hur heterogena elevgrupperna är och att elevers skolframgång hänger nära samman med elevernas tidigare utbildningsbakgrund. Lärarna i den aktuella studien beskrev att de i genomsnitt har 3 till 5 olika språk i sina klasser, representerade bland ett litet antal elever. Antalet språk visade sig vara oberoende av skolform medan det totala antalet nyanlända skiftar, på språkintröfning är alla nyanlända eller språkinlärare. När det gäller lärarnas erfarenhet av dessa elever visade det sig att en majoritet hade 6 års erfarenhet eller mer (2018 då enkäten genomfördes), vilket relaterar tillbaka till tiden före 2015 då den stora vågen nyanlända kom. Frekvensen för hur ofta lärarna haft språkinlärare i sina grupper varierade dock mellan "ibland", "då och då", respektive "hela tiden".

Det som angavs vara svårast för eleverna är långa skrivuppgifter, vilket kan anses föga förvånande. Däremot menade lärarna att de uppfattar korta skrivuppgifter som enklare för eleverna än förberedda muntliga uppgifter. Vid frågan om vilka språkliga register som lärarna uppfattar vara svårast angav en majoritet ämnesspecifika begrepp, vilket överensstämmer med vad de fokuserar på i undervisningen som beskrivits ovan. På frågan om de samarbetar med en kollega och i så fall med vem och vad, angav de flesta att de samarbetar med ämneskollegor. De hjälps då framför allt åt med utformning av bedömningsuppgifter och granskning av elevprestationer. Ämneskollegor var även de som lärarna angav att de vänder sig till för att få idéer hur de kan arbeta med ämnesspråk och allmänt akademiskt språk. Om man samarbetade med någon annan var det framför allt lärare i svenska som andraspråk.

Lärare var påfallande eniga om att bedömning ska ske på svenska, och hänvisade framför allt till en outtalad policy. De flesta ansåg dock att det inte finns någon policy avseende språkanvändning. Mycket få rapporterade att de använder sig av elevers starkaste språk vid bedömning, vilket rekommenderas i råd för bedömning av nyanlända åtminstone initialt (National Agency for Education, 2016). I intervjuerna påpekade lärarna svårigheterna som skulle vara förenade med detta då de inte behärskar elevernas språk själva. De var även tveksamma till att använda sig av översättningar och modersmålslärare då det innebär att kontrollen över vad elever kan flyttas längre från dem själva. Vissa lärare påtalade även det orättvisa i att låta de elever som har tillgång till en lärare som kan deras språk använda det medan andra inte har samma möjlighet då deras språk inte finns representerat på skolan.

Den vanligaste formen av stöttning rapporterades vara bilder eller visuellt stöd och användning av förenklat språk. Minst vanligt förekommande var att uppgifter översätts till elevens språk eller att ha uppgiftsformat som erbjuder start-meningar. Vid nationella prov, om elever genomför dessa, angavs förlängd provtid vara den vanligaste anpassningen. En fjärdedel rapporterade att de låter elever göra proven på engelska, vilket kan sägas utgöra ett hot mot validiteten på samma vis som om elever får göra prov på svenska i CLIL, som beskrevs i Studie 1 ovan.

I den sista forskningsfrågan efterfrågades om lärares syn på bedömning och bedömningspraktik påverkas av skolform och ämne och i så fall hur och på vilka grunder. I enkäten framgick att lärare med en dubbel ämnesbehörighet i ett språkämne och ett samhällsvetenskapligt ämne visade statistiskt signifikant högre språklig fokus med användning av modell texter i bedömningsuppgifter och förberedelser som inkluderar genomgång av användbara fraser, studier av ämnestexter, skriva texter tillsammans i klassrummet och att förse elever med exempeltexter och illustrerande modeller för språkanvändning. De lärare i som i intervjuer sade att de använder sig av liknande, både inom natur- och samhällsvetenskapliga ämnen, angav också att de ser bäst progression hos sina elever, även om utmaningar ofta kvarstår av olika skäl. I jämförelsen mellan skolformer framgick att studier av

ämnestexter var statistiskt signifikant högre på språkintröduktion än i grundskolans årskurs 7-9, medan den uppskattade svårighetsgraden i alla aktiviteter, muntliga såväl som skriftliga, var statistiskt signifikant högre i årskurs 7-9. Det kan anses ligga i sakens natur att elever som når gymnasienivå har något lättare för dessa aktiviteter om man antar att de inte är helt nyanlända och även uppnått större mognad. Att argumentera ansågs därmed lättare och lättare för varje skolnivå. Det kan även ha att göra med pedagogiken, lärare på språkintröduktion ansåg exempelvis att ämnesspecifika begrepp var minst svårt för eleverna av de tre lärarkategorierna och möjligen arbetar de mer med detta, men inga statistiskt signifikanta resultat noterades för lärare i någon skolform avseende arbetet med begrepp. Vad gäller återkoppling på språket skiljde gymnasielärarna ut sig som de som ger mest korriigerande återkoppling medan transspråkande och användning av elevers starkaste språk angavs vara mer frekvent på högstadiet.

Diskussion och syntes

De två övergripande forskningsfrågorna förenar och jämför frågor rörande förhållningssätt till undervisningsspråket i ämnesmålen, vid bedömning och i undervisning i de två kontexterna CLIL och NAS. Finns det skillnader och/eller gemensamma drag?

När det gäller lärares syn på bedömning framkom framför allt många gemensamma drag. I båda kontexterna låg fokus på kunskapskraven i kurs- och ämnesplanerna, vilket kan benämnas som en bedömarkultur, på engelska "accountability culture". Även om lärare bekräftade elevers svårigheter att visa sitt ämneskunnande till fullo, återkom de till vad som stipuleras i målen. I CLIL sade lärarna uttryckligen att eleverna förmodligen får acceptera ett något lägre betyg än vad de skulle ha fått om de gjort kursen på svenska, vilket de sade att de är tydliga med till eleverna i början av deras gymnasietid. En av historielärarna påpekade att det är svårt att visa tillräcklig ämnesförståelse på engelska vilket i historia ofta handlar om att kunna utveckla sina svar i längre texter för att visa sin analytiska eller berättande förmåga. Bland NAS beskrev lärare på motsvarande vis att eleverna förmodligen får sikta på godkänt betyg, E, men att det förekommer att elever får högre betyg, C och även B. Detta gällde särskilt elever som integreras snabbt enligt lärarna och snabbt kommer in i svenska språket.

Både i CLIL och bland NAS sade lärarna att de skriftliga proven tillmätts särskild vikt, delvis för att eleverna själva föredrar skriftliga uppgifter. Lärare i båda sammanhangen angav att de använder sig av enskilda muntliga uppföljningar då eleverna inte lyckas visa sitt ämneskunnande i skrift. Bland NAS rapporterade lärarna att de ibland används som alternativ till en skrivuppgift. I den sistnämnda kontexten rapporterade flera av de intervjuade lärarna, framför allt i årskurserna 7-9 och på introduktionsprogrammet, att de helst skulle använda bara muntliga uppgifter om de fick. I enkäten skrev många att de föredrar att blanda uppgiftsformat utifrån vad som fungerar bäst för eleven, vilket visar på en elevcentrerad pedagogik.

Både CLIL och NAS-lärares bedömningsformer visade sig följa ämnets karaktär och ämnesgenre. Skriftliga prov ansågs viktigt i naturvetenskapliga ämnen, genre definierades bland annat som laborationsrapporter. I samhällsvetenskapliga discipliner fästes mer vikt vid längre resonerande utredande frågor och texter. I Studie I beskrev engelsklärarna på liknande vis att de gärna arbetar med elevers skrivande, och då i olika textformat. Detta skulle därmed kunna vara ett område för ämnesövergripande samarbete. Engelsklärarna i Studie I sade att de inte hinner ge formativ återkoppling till alla elever, vilket de beskrev som orimligt och menade även att de inte skulle kunna säga något om innehållet i fall elever skulle skriva om något ämne de inte är bekanta med, från exempelvis biologisk.

Lärares förhållningssätt till undervisningsspråket i de två kontexterna framstod som lite kluvet. I den första studien hävdade ämneslärare såväl som engelsklärare att det inte är ämneslärares roll att bedöma elevernas engelska, lika lite som det skulle vara i svenska, enligt en av lärarna. I CLIL hävdade dessutom vissa av ämneslärarna att eleverna kan välja att använda ämneskoncept på engelska eller svenska så länge de gör sig förstådda, vilket kan ses som en form av transspråkande eller en stöttning. Det skapar dock vissa validitetsfrågor utifrån bristen på samband mellan undervisningsspråk och språk vid bedömning som nämndes i teoridelen (Abedi, 2009; Gablasova, 2014). Lärarna beskrev ändå att progression i ämneskunskan hänger på elevers förmåga att förklara och jämföra, resonera och dra slutsatser exempelvis. I Studie I var det samtidigt enbart engelsklärare som angav att det är viktigt med allmänt akademiskt språk vid bedömning i ämnet. I Studie II beskrev flera lärare i olika ämnesdiscipliner hur de stöttar eleverna språkligt och att detta tillsammans med om eleverna inte använder sitt modersmål utan snabbt kommer in i svenska, har en positiv effekt på elevernas framgång i ämnet.

I CLIL uttryckte lärare sig lite olika huruvida de anser att enbart engelska ska användas eller om läraren tillåter transspråkande eller användande av svenska. Bland NAS var lärare relativt eniga att det är svenska som gäller. Några uttryckte sig lite undrande över transspråkande då det fördes på tal, men att det kunde vara intressant att veta mer om, medan ett par grundskollärare beskrev att de använder sig av det. Samtliga hävdade dock att det kan fungera initialt eller på språkintröskning, men på gymnasiet behöver eleverna behärska svenska för att exempelvis kunna skriva längre texter enligt lärarna i alla skolformer.

I Studie II uttryckte lärarna att de helst vill samarbeta med andra ämneskollegor då andra lärare inte behärskar deras ämnen, vilket belyser behovet att hitta tydliga former för hur ett sådant samarbete i så fall skulle kunna utformas för att samordna lärares olika kompetenser. Ett alternativ är annars att ämneslärare blir mer av språklärare själva, vilket också innebär behov av en tydlig språkpedagogik. Som förslag på utvecklingsbehov angav lärare att de ville ha tydligare och mer uttryckliga riktlinjer för bedömning och att anpassa kursplanen i samhällsämnen efter kursplanen i svenska som andraspråk för att hitta en "lightvariant" för nyanlända. De efterfrågade bedömda elevtexter för E för att kunna visa på vilken nivå eleverna förväntas ligga för godkänt betyg, och tydliga råd hur de ska göra för att bedöma ämneskunskan snarare än svenska.

Majoriteten av lärarna verkade ha en stor medvetenhet om vad de ska bedöma och hur detta lämpligast görs med utgångspunkt i ämnets karaktär. Anpassningar var vanligare bland nyanlända, men även i CLIL såg lärarna ett behov att hitta lösningar för att komma förbi språkliga hinder. Bristande språk fick med andra ord konsekvenser för bedömningen i ämnena. Utifrån de modeller som beskrevs som en del av det teoretiska ramverket och som användes som analysverktyg verkade en majoritet av lärarna ha en relativt osynlig språkpedagogik där fokus framför allt låg på ämnesbegrepp och inte så mycket på undervisning av ämnesövergripande språk. Betoningen på kunskapskrav och förmågor visade även på en stor grad av central styrning, vilket ligger rätt långt ifrån transspråkande-pedagogik. Länkningen mellan vad som förväntas uppnås utifrån kunskapskraven och den implementerade läroplanen var oftast inte koherent utifrån ett andraspråksinläraresperspektiv. Goda intentioner verkade finnas vilka beskrevs som enskild uppföljning bland annat, men flertalet lärare rapporterade att de inte fokuserade på undervisning av allmänt akademiskt språk eller ämnesspråk utöver begrepp. Något fler lärare i Studie II jämfört med Studie I angav dock att de börjat bli mer "språklärare". För att undvika konstrukt underrepresentation och avvärja hotet mot icke valida bedömningar behöver elever ges tillfälle att lära sig de olika språkliga element som ingår i en ämnesgenre som akademisk argumenterande uppsats eller laborationsrapport. Det handlar om länkning mellan förväntade kursmål, undervisning och de bedömningsformer som ska användas.

Implikationer

Med dessa två studier hoppas jag ha kunnat belysa behovet att explicitgöra vem som har ansvaret för elevers språkutveckling i ämnena och därmed gör ämnesstoff tillgängligt för alla elever så att kunskapsbedömningar kan bli så valida som möjligt. Hur språket behandlas i ämnena beskriver jag i avhandlingen som antingen "language by default" eller "language by design". Det är endast det sistnämnda fallet som präglas av en genomtänkt koherent strategi (jfr "konstruktiv länkning") medan det första innebär ett överhängande hot om icke jämförbara och ojämlika förhållanden, resultat och till sist betyg. Language by design sätter fokus på att lära ut nödvändigt språk som behövs för att kunna utföra förväntade språkliga handlingar bland annat i samband med bedömningssituationer.

Utifrån mina resultat skulle jag även vilja väcka en diskussion om behovet av en enhetlig språkpolicy över olika undervisningskontexter i svensk skola i relation till målspråk och undervisningsspråk. I svensk skola förekommer både en uttalad och outtalad förväntan att elever ska lära sig svenska, men i CLIL-sammanhang kan elever lära sig vissa ämnesregister på enbart engelska. Bland nyanlända hänvisas till transspråkande och användning av elevers starkaste språk, men det är beroende av tillgång till lärare med den språkliga kompetensen vilket varierar. Likaså saknas beskrivningar av hur en transspråkspraktik ska infogas i utvecklingen mot målspråket svenska. I båda kontexterna, bland nyanlända och i CLIL, finns en motsättning i förhållande till språklig korrekthet. Ämneslärare anger att det inte ser sig som språklärare och därmed kan de inte korrigera språket, i CLIL hänvisar de dessutom till att de själva är andraspråkstalare. Vad får allt detta för konsekvenser? I forskning och utredning av CLIL i svensk kontext (National Agency for Education, 2018h; Sylvé, 2019) noteras att Sveriges avsaknad av riktlinjer och policy för språkanvändning då engelska är undervisningsspråk är unikt i internationell jämförelse. Många andra länder har krav på relevant lärarutbildning eller har utvecklat specifika språkmål i ämnena. I utbildning av andraspråkelever talas det om behovet av att synliggöra olika språkliga register i ämnesundervisningen och att ha en uttalad ansvarsfördelning mellan ämneslärare och språklärare alternativt samarbete (Cummins, 2017; Schleppegrell, 2001, 2004, 2020). Detta för en hållbar utveckling i samhället i stort, men till att börja med för rättvisa och jämförbara resultat för alla elever i flerspråkig skola och utbildning (Kane, 2010; OECD, 2009b, 2015a, 2015b; Stobart, 2005).

Tre huvudspår med förslag till utveckling för att förbättra utbildningen för NAS framträder som ett resultat av den här studien och de berör a) lärarens utbildning och kompetensutveckling, b) organisationen på skolorna och c) kurs- och ämnesplanerna. Lärare ger uttryck för dessa antingen explicit eller implicit:

- Ämneslärare behöver utveckla språklärarkompetenser och se sig som språklärare inom sina egna ämnen. Kompetensutveckling för lärare och lärarutbildning behöver fortsätta att fokusera på hur ämnesrelaterat språk liksom övergripande skolspråk synliggörs i undervisning.
- Ämnesövergripande samarbete behöver vara norm. Gemensamma riktlinjer och en organisation behöver finnas på plats som underlättar och främjar samarbete med tydlig rollfördelning.
- Reviderade kurs- och ämnesplaner med SvA-inriktning. Renodla det viktigaste innehållet, formulera språkliga mål och ge elevexempel på olika språkliga nivåer: särskilt för E-nivå på svenska för NAS.

Alla dessa tre punkter går att överföra på CLIL-undervisning för att innebära "language by design" snarare än "by default". Även om det inte skulle kallas SvA-inriktning i CLIL, för att eleverna där kan anses som mer högpresterande och gynnade, kvarstår behovet av tydliga språkliga mål och språkpolicy.

Det råder brist på samordning mellan språk i undervisning och bedömning respektive olika krav på användning av målspråket i CLIL jämfört med NAS, och hur omsätts transspråkande i praktiken och vad är målet? Vad är konsekvenserna av att ha tillägnat sig ett ”hyfsat” akademiskt språk i engelska av en svensktalande lärare, möjligen med ett lägre ämnesbetyg, alternativt en hyfsad svenska där jag gör mig förstådd men lärare inte vill rätta för att det kan uppfattas nedvärderande? En av lärarna, själv med invandrarbakgrund, var den som uttryckte det allra tydligast och det avslutar den här sammanfattningen:

Svenska elever, oavsett etnicitet, har inte medvetande om klasstillhörighet. Dom är privilegierade hela vägen upp [...] men språket är väldigt viktigt och det finns minst två Sverige när det gäller språk och det svåra är också att ingen kommer att påtala dom språkliga fel som identifierar någon som lägre klass om inte läraren eller omgivningen gör detta väldigt tidigt, så jag tycker att jag ska göra det.

Syftet med föreliggande studie var att bidra till forskning om bedömning i flerspråkig miljö bland språkinlärare som får sitt ämneskunnande bedömt på ett andraspråk/främmande språk. För att uppnå en så rättvis och valid bedömning som möjligt är det av avgörande betydelse att alla elever får jämförbara förutsättningar att inhämta både kunskap och förmågor där mål, undervisning och bedömningsformer hänger samman. Bedömning i en språkligt heterogen miljö behöver mer forsknings- och erfarenhetsbaserad kunskap. Mycket kunnande finns att hämta hos lärare och elever med erfarenhet från detta relativt obeforskade område och förhoppningen är att jag med denna studie lyckats bidra med en liten del och att mer forskning kommer.

References

Note: Scandinavian authors are listed according to the English alphabet (e.g. Ö = O and Ø = O).

- Abedi, J. (2008). Utilizing Accommodations in Assessment. In E. Shohamy & N. H. Hornberger (Eds.), *Language Testing and Assessment* (Vol. 7, pp. 331–347).
- Abedi, J. (2009). Validity of Assessments for English Language Learning Students in a National/ International Context. *Estudios Sobre Educación*, 16, 167–183.
- Abell, S. K. (2007). Research on science teacher knowledge. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (Handbook of research on science education, pp. 1105–1149). Lawrence Erlbaum Associates.
- Abell, S. K., & Siegel, M. A. (2011). Assessment Literacy: What Science Teachers Need to Know and Be Able to Do. In D. Corrigan, J. Dillon, & R. Gunstone (Eds.), *The Professional Knowledge Base of Science Teaching* (pp. 205–221). Springer Netherlands. https://doi.org/10.1007/978-90-481-3927-9_12
- Ahn, E. (2007). *Conference paper prepared for the European Union Studies Association 2007 Biennial Conference (Montreal, Canada)*. 26.
- Aikenhead, G. S. (2005). *Science Education for Everyday Life*. Teachers' College Press.
- Airey, J. (2012). "I don't teach language": The linguistic attitudes of physics lecturers in Sweden. *AILA Review*, 25, 64–79. <https://doi.org/10.1075/aila.25.05air>
- Akker, J. V. D. (1997). The science curriculum: Between ideals and outcomes. *The International Handbook of Science Education*, 421–447.
- Akker, J. V. D. (1999). Principles and methods of development research. In K. Gustafson, J. V. den Akker, M. Nieveen, & T. Plomp (Eds.), *Design approaches and tools in education and training*. (pp. 1–14). Kluwer Academic Publishers.
- Allwood, C. M. (2007). Otydlig gräns mellan kvalitativ och kvantitativ forskning. [Unclear boundary between qualitative and quantitative research]. *Psykologitidningen*, 4(07), 25–27.
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and, & Psychological Testing (U.S.). (1999). *Standards for educational and psychological testing*. AERA.
- Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives: Complete Edition*. Longman.
- Andersson, A., Lyrenäs, S., & Sidenhag, L. (2015). *Nyanlända i skolan. [Newly arrived students in school.]*. Gleerups utbildning.
- Andersson, B., Bach, F., Olander, C., & Zetterqvist, A. (2005). *Naturorienterande ämnen. Nationella Utvärderingen av grundskolan 2003. (NU-03) [Natural science subjects: The National Evaluation of Compulsory School 2003.]*. Skolverket.
- Anker-Hansen, J. (2015). *Assessing scientific literacy as participation in civic practices affordances and constraints for developing a practice for authentic classroom assessment of argumentation, source critique and decision-making*. [Doctoral dissertation] Department of Mathematics and Science Education, Stockholm University. <http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-119866>
- Areskoug, M., Ekborg, M., Lindahl, B., & Rosberg, M. (2013). *Naturvetenskapens bärande idéer: För lärare 1-6. [The foundational principles in the natural sciences: For teachers in grades 1-6.]*. Gleerup.

- Aronin, D. L., & Singleton, D. (2008). Multilingualism as a New Linguistic Dispensation. *International Journal of Multilingualism*, 5(1), 1–16. <https://doi.org/10.2167/ijm072.0>
- Åsberg, R., Hummerdal, D., & Dekker, S. (2011). There are no qualitative methods – nor quantitative for that matter: The misleading rhetoric of the qualitative–quantitative argument. *Theoretical Issues in Ergonomics Science*, 12(5), 408–415. <https://doi.org/10.1080/1464536X.2011.559292>
- Axelsson, M., Olofsson, M., Philipsson, A., Rosander, C., & Sellgren, M. (2006). *Ämne och språk—Språkliga dimensioner i ämnesundervisningen. [Subject and language—Language dimensions in content instruction.]*. Kompetensfonden.
- Bachman, L. F., & Palmer, Adrian. S. (2010). *Language assessment in Practice*. Oxford University Press.
- Baker, E. L., O’ Neil, H. F., & Linn, R. L. (1993). Policy and validity prospects for performance-based assessment—ProQuest. *American Psychologist*, 48(12), 1210–1218. <https://doi.org/10.1037/0003-066X.48.12.1210>
- Bakhtin, M. M. (1981). *The Dialogic Imagination: Four Essays* (M. Holquist, Ed.; C. Emerson & M. Holquist, Trans.). University of Texas Press.
- Bal, A., & Trainor, A. A. (2016). Culturally Responsive Experimental Intervention Studies: The Development of a Rubric for Paradigm Expansion. *Review of Educational Research*, 86(2), 319–359. <https://doi.org/10.3102/0034654315585004>
- Ball, P. (n.d.). *Article: How do you know if you’re practising CLIL?* Onestopenglish. Retrieved May 26, 2020, from <http://www.onestopenglish.com/clil/methodology/articles/article-how-do-you-know-if-youre-practising-clil/500614.article>
- Ball, P., Kelly, K., & Clegg, J. (2016). *Putting CLIL into Practice: Oxford Handbooks for Language Teachers*. Oxford University Press.
- Banks, J. A. (2004). Teaching for Social Justice, Diversity, and Citizenship in a Global World. *The Educational Forum*, 68(4), 296–305. <https://doi.org/10.1080/00131720408984645>
- Banks, J. A. (2009). Multicultural education: Dimensions and paradigms. In J. A. Banks (Ed.), *The Routledge international companion to multicultural education*. (pp. 9–33). Routledge.
- Barnes, L. L. B., Bull, K. S., Campbell, N. J., & Perry, K. M. (2001). Effects of Academic Discipline and Teaching Goals in Predicting Grading Beliefs Among Undergraduate Teaching Faculty. *Research in Higher Education*, 42(4), 13.
- Barnes, N., Fives, H., & Dacey, C. M. (2015). *Teacher Beliefs about Assessment*. Montclair State University, US.
- Barwell, R. (2014). Centripetal and centrifugal language forces in one elementary school second language mathematics classroom. *ZDM*, 46(6), 911–922.
- Beacco, J.-C., Fleming, M., & Goullier, F. (2016). *The language dimension in all subjects: A handbook for curriculum development and teacher training*. Council of Europe.
- Bell, L., & Stevenson, H. (2006). *Education policy: Process, themes and impact*. Routledge. <http://eprints.lincoln.ac.uk/id/eprint/1851/>
- Bergendorff, I. (2014). Nyanlända elever i Sverige. [Newly arrived students in Sweden.]. In G. Kästen-Ebeling & T. Otterup (Eds.), *En bra början—Mottagande och introduktion av nyanlända elever. [A good start—Reception and introduction of newly arrived students.]* (pp. 31–48). Studentlitteratur.
- Berger, J.-L., Girardet, C., Vaudroz, C., & Crahay, M. (2018). Teaching Experience, Teachers’ Beliefs, and Self-Reported Classroom Management Practices: A Coherent Network. *SAGE Open*, 8(1), 215824401775411. <https://doi.org/10.1177/2158244017754119>
- Berger, R. (2015). Now I See it, Now I Don’t: Researcher’s Position and Reflexivity in Qualitative Research. *Qualitative Research*, 15(2), 219–234.

- Berglund, J. (2017). Education Policy – A Swedish Success Story? *EDUCATION POLICY*, 12.
- Bernstein, B. (1999). Vertical and Horizontal Discourse: An essay. *British Journal of Sociology of Education*, 20(2), 157–173. <https://doi.org/10.1080/01425699995380>
- Bernstein, B. (2000). *Pedagogy, Symbolic Control and Identity: Theory, Research, Critique* (2nd ed.). Rowman & Littlefield.
- Biggs, J. (2003). *Aligning teaching and assessment to curriculum objectives*. LTSN Generic Centre.
- Biggs, J., & Tang, C. S. (1999). *Teaching for Quality Learning at University*.
- Bjerregaard, M., & Kindenberg, B. (2015). *Språkutvecklande SO-undervisning—Strategier och metoder för högstadiet. [A language development perspective in the teaching of the social sciences—Strategies and methods for secondary education.].* Natur & Kultur.
- Black, P., & Wiliam, D. (1998). *Inside the Black Box: Raising standards through classroom assessment*. School of Education, King's College.
- Bloom, B., S. (1956). *Major categories in the cognitive domain of the taxonomy of educational objectives*.
- Bloom, B., S., Engelhart, M., D., Furst, E., J., Hill, W., H., & Krathwohl, D., R. (1956). *Taxonomy of Educational Objectives. The Classification of Educational Goals*. Longmans Green and co. <https://www.uky.edu/~rsand1/china2018/texts/Bloom%20et%20al%20-Taxonomy%20of%20Educational%20Objectives.pdf>
- Bolton, K., & Kuteeva, M. (2009). *The Survey of the Use of English at Stockholm University* (p. 24). Stockholm University.
- Bonjean, D. (2018, August 6). *About multilingualism policy* [Text]. Education and Training - European Commission. https://ec.europa.eu/education/policies/multilingualism/about-multilingualism-policy_en
- Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, 36, 81–109.
- Brandén, M., & Bygren, M. (2018). *School Choice and School Segregation: Lessons from Sweden's School Voucher System*. 43.
- Braun, V., Clarke, V., Hayfield, N., & Terry, G. (2019). Thematic Analysis. In P. Liamputtong (Ed.), *Handbook of Research Methods in Health Social Sciences* (pp. 843–860). Springer. https://doi.org/10.1007/978-981-10-5251-4_103
- Bray, M., Adamson, R. B., & Mason, M. (Eds.). (2007). *Comparative Education Research: Approaches and Methods*. Springer Netherlands. <https://doi.org/10.1007/978-1-4020-6189-9>
- Bridges, D. (2002). The Ethics of Outsider Research. In M. McNamee & D. Bridges (Eds.), *The Ethics of Educational Research*. Blackwell Publishers Ltd.
- Brooks, L. (2009). Interacting in pairs in a test of oral proficiency: Co-constructing a better performance. *Language Testing*, 26(3), 341–366. <https://doi.org/10.1177/0265532209104666>
- Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11, 301–318.
- Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged instrument. *Psychological Reports*, 99, 166–170.
- Brown, G. T. L. (2008). *Conceptions of Assessment: Understanding What Assessment Means to Teachers and Students*. Nova Science Publishers. https://www.researchgate.net/publication/233871197_Conceptions_of_Assessment_Understanding_What_Assessment_Means_to_Teachers_and_Students
- Brown, J. D., & Hudson, T. (2002). *Criterion-Referenced Language Testing*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139524803>
- Bryman, A. (2016). *Samhällsvetenskapliga metoder. [Social Research Methods]* (3rd ed.). Liber.
- Bryman, A., & Cramer, D. (1990). *Quantitative Data Analysis for Social Scientists*. Routledge.

- Bulterman-Bos, J. A. (2008). Will a Clinical Approach Make Education Research More Relevant for Practice? *Educational Researcher*, 37(7), 412–420.
<https://doi.org/10.3102/0013189X08325555>
- Bunar, N. (2010). *Nyanlända och lärande: En forskningsöversikt om nyanlända elever i den svenska skolan. [Newly-arrived students and learning: a research review of newly arrived students in Swedish schools.]*. Vetenskapsrådet.
- Busch, B. (2012). The Linguistic Repertoire Revisited. *Applied Linguistics*, 33(5), 503–523.
<https://doi.org/10.1093/applin/ams056>
- Cabau-Lampa, B. (1999). Decisive factors for language teaching in Sweden. *Educational Studies*, 25, 175–186.
- Callaghan, M., Knapp, P., & Noble, G. (1993). Genre in Practice. In B. Cope & M. Kalantzis (Eds.), *The Powers of Literacy: A Genre Approach to Teaching Writing*. University of Pittsburgh Press.
- Camara, W. J. (2015). *Standards for Educational and Psychological Testing*. 15.
- Carlgren, I. (2013). Kunnande—Kunskap—Kunnighet. In L. Lindström, V. Lindberg, & A. Pettersson (Eds.), *Pedagogisk bedömning. Att dokumentera, bedöma och utveckla kunskap*. (3rd ed.). Liber.
- Casservik, M. (2005). Vem styr?: Om lärares planering av historieundervisning. In *Historiedidaktiska perspektiv: Bidrag från lärare och studenter vid lärarutbildningen i Jönköping*. (pp. 265–324). Jönköping University Press.
- Cederberg, M. (2012). *Gymnasieskolan- inte en skola för alla? En forskningssammanställning om låg utbildning och hälsa. [Upper secondary school—Not a school for everyone? A research survey about low skills and poor health.]*. Kommission för ett socialt hållbart Malmö.
- Cenoz, J. (2013). Defining Multilingualism. *Annual Review of Applied Linguistics*, 33, 3–18.
<https://doi.org/10.1017/S026719051300007X>
- Cenoz, J. (2015). Content-based instruction and content and language integrated learning: The same or different? *Language, Culture and Curriculum*, 28(1), 8–24.
<https://doi.org/10.1080/07908318.2014.1000922>
- Chalhoub-Deville, M. (2016). Validity theory: Reform policies, accountability testing, and consequences. *Language Testing*, 33(4), 453–472. <https://doi.org/10.1177/0265532215593312>
- Cohen, L., Manion, L., & Morrison, K. (2011). *Research Methods in Education* (7th ed.). Routledge.
- Conteh, J., & Meier, G. (Eds.). (2014). *The Multilingual Turn in Languages Education: Opportunities and Challenges*. (Vol. 40). Multilingual Matters.
- Corrigan, D., Dillon, J., & Gunstone, R. (Eds.). (2011). *The Professional Knowledge Base of Science Teaching*. Springer.
- Council of Europe (n.d.). *Council of Europe Language Policy Portal*. Retrieved April 30, 2020, from <https://www.coe.int/en/web/language-policy>
- Council of Europe (Ed.). (2011). *The Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)*. <https://rm.coe.int/1680459f97>
- Coyle, D. (1999). Supporting students in content and language integrated learning contexts: Planning for effective classrooms. In J. Masih (Ed.), *Learning through a foreign language. Models, methods and outcomes*. CILT.
- Coyle, D. (2005). *Planning and Monitoring CLIL*. 17.
- Coyle, D., Holmes, B., & King, L. (2009). *Towards an integrated curriculum – CLIL National Statement and Guidelines*. The Languages Company.
- Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and Language Integrated Learning*. Cambridge University Press.

- Crohnbach, L. J. (1971). Test Validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.) (pp. 443–507). American Council on Education.
- Crooks, T. J., Kane, M. T., & Cohen, A. S. (1996). Threats to the Valid Use of Assessments. *Assessment in Education: Principles, Policy & Practice*, 3(3), 265–286. <https://doi.org/10.1080/0969594960030302>
- Cummins, J. (1979). *Cognitive/Academic Language Proficiency, Linguistic Interdependence, the Optimum Age Question and Some Other Matters. Working Papers on Bilingualism*, No. 19. Bilingual Education Project, The Ontario Institute for Studies in Education, 252 Bloor St. <https://eric.ed.gov/?id=ED184334>
- Cummins, J. (1981). The Role of Primary Language Development in Promoting Educational Success for Language Minority Students. In California State Department of Education (Ed.), *Schooling and language minority students: A theoretical framework*. California state Department of Education.
- Cummins, J. (1984). *Bilingualism and Special Education: Issues in Assessment and Pedagogy*. Multilingual Matters.
- Cummins, J. (2000). *Language, Power and Pedagogy. Bilingual Children in the Crossfire*. Multilingual Matters.
- Cummins, J. (2008). BICS and CALP: Empirical and Theoretical Status of the Distinction. In B. Street & N. H. Hornberger (Eds.), *Encyclopedia of Language and Education, Volume 2: Literacy* (pp. 487–499). Springer US. https://doi.org/10.1007/978-0-387-30424-3_36
- Cummins, J. (2009). Fundamental psycholinguistic and sociological principles underlying educational success for linguistic minority students. In T. Skutnabb-Kangas, R. Phillipson, A. K. Mohanty, & M. Panda (Eds.), *Social justice through multilingual education*. (pp. 19–36). Multilingual Matters.
- Cummins, J. (2017). *Flerspråkiga elever. Effektiv undervisning i en utmanande tid. [Multilingual students. Effective teaching in challenging times.]*. Natur & Kultur.
- Cummins, J., Brown, K., & Sayers, D. (2006). *Literacy, Technology, and Diversity: Teaching for Success in Changing Times* (PAP/CDR edition). Pearson.
- Czura, A., & Papaja, K. (2013). Curricular models of CLIL education in Poland. *International Journal of Bilingual Education and Bilingualism*, 16(3), 321–333. <https://doi.org/10.1080/13670050.2013.777388>
- Dagens Arena. (2019, December 12). Nytt mål i lärarutbildningen skulle göra skillnad för PISA. [A new goal in teacher education would make a difference for PISA.]. *Dagens Arena*. <https://www.dagensarena.se/opinion/nytt-mal-lararutbildningen-skulle-gora-skillnad-pisa/>
- Dagens samhälle. (2017, April 20). Här är de mest segregerade skolorna. [Here are the most segregated schools.]. <https://www.dagenssamhalle.se/nyhet/haer-aer-de-mest-segregerade-skolorna-33206>
- Dalberg, T. (2013). *Engelskans utbredning i den svenska högskolan. En kartläggning av engelskspråkiga benämningar av program och kurser i svensk högre utbildning 1993-2009. [The spread of English in Swedish Universities. Mapping of English program- and course descriptions in Swedish tertiary education.]* (No. 52; Sociology of Education and Culture Research Reports). SEC, Uppsala universitet. <https://uu.diva-portal.org/smash/get/diva2:667893/FULLTEXT01.pdf>
- Dalton-Puffer, C. (2007). *Disocurse in Content and Language Integrated Learning (CLIL) Classrooms*. John Benjamins Publishing Company.
- Dalton-Puffer, C. (2016). Cognitive Discourse Functions: Specifying an Integrative Interdisciplinary Construct. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualising Integration in CLIL and Multilingual Education* (pp. 29–54). Multilingual Matters.

- de Silva Joyce, H., & Burns, A. (1999). *Focus on Grammar*. Macquarie University, National Centre for English Language Teaching and Research.
- Denzin, N. K., & Lincoln, Y. S. (Eds.). (1994). *Handbook of qualitative research*. SAGE.
- Department of Education. (2015). [Utbildningsdepartementet] u2015-03844-Uppdrag om nationella skolutvecklingsprogram. [Mission for national school development programs.].
<https://www.regeringen.se/4a2aff/globalassets/regeringen/dokument/utbildningsdepartementet/barn--och-ungdomsutbildning/u2015-03844-s-uppdrag-om-nationella-skolutvecklingsprogram.pdf>
- Ducasse, A. M., & Brown, A. (2009). Assessing paired orals: Raters' orientation to interaction: *Language Testing*. <https://doi.org/10.1177/0265532209104669>
- Economic Policy Institute. (2018). *2018 Environmental Performance Index. Global metrics for the environment: Ranking country performance on high-priority environmental issues*. Yale University. <https://epi.envirocenter.yale.edu/downloads/epi2018policymakerssummaryv01.pdf>
- Educational Testing Service. (2014). *E. T. S. Standards for Quality and Fairness 2014*. 70.
- Educational Testing Service. (2016). *ETS Guidelines for Fair Tests and Communications*. Educational Testing Service.
- Ekström, A. (2020, January 31). Komvux ändras för att stärka integration och kompetens. [Reforms in adult education to strengthen integration and competence.]. *Dagens Nyheter*.
<https://www.regeringen.se/debattartiklar/2020/01/komvux-andras-for-att-starka-integration-och-kompetens/>
- Elmeroth, E. (2017). *Möte med andraspråkselever. [Meeting second language learners.]*. Studentlitteratur.
- Entwistle, N., Skinner, D., Entwistle, D., & Orr, S. (2000). Conceptions and beliefs about good teaching: An integration of contrasting research areas. *Higher Education Research & Development*, 19(1), 5–26.
- Erickson, G., & Gustafsson, J.-E. (2017). Bedömningens dubbla funktion—För lärande och likvärdighet. [The twofold function of assessment—For learning and equity.]. In U. P. Lundgren, R. Säljö, & C. Liberg (Eds.), *Lärande, skola, bildning. Grundbok för lärare*. (4th ed.). Natur & Kultur.
- European Commission. (2003). *COMMUNICATION FROM THE COMMISSION TO THE COUNCIL, THE EUROPEAN PARLIAMENT, THE ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Promoting Language Learning and Linguistic Diversity: An Action Plan 2004 – 2006*. <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2003:0449:FIN:EN:PDF>
- European Commission. (2019). *Education and training monitor 2019—Sweden*. 24.
- European Council. (2002). *Presidency conclusions—Barcelona 15 and 16 March 2002*.
[Http://Aei.Pitt.Edu/43345/1/Barcelona_2002_1.Pdf](http://Aei.Pitt.Edu/43345/1/Barcelona_2002_1.Pdf).
http://aei.pitt.edu/43345/1/Barcelona_2002_1.pdf
- Eurostat. (2018). *65% know at least one foreign language in the EU*.
<https://ec.europa.eu/eurostat/web/products-eurostat-news/-/EDN-20180926-1>
- Eurydice. (2006). *Content and Language Integrated Learning (CLIL) at school in Europe*. European Commission. <http://op.europa.eu/en/publication-detail/-/publication/756ebdaa-f694-44e4-8409-21eef02c9b9b>
- Eurydice. (2017). *New Eurydice Brief Published: Key Data on Teaching Languages at School in Europe, 2017 Edition*. https://eacea.ec.europa.eu/erasmus-plus/news/new-eurydice-brief-published-key-data-teaching-languages-school-in-europe-2017-edition_en
- Examples of Applied Research Topics*. (n.d.). YourDictionary. Retrieved March 28, 2020, from <https://examples.yourdictionary.com/examples-of-applied-research.html>

- Fives, H., & Buehl, M. M. (2016). Teachers' Beliefs, in the Context of Policy Reform. *Policy Insights from the Behavioral and Brain Sciences*, 3(1), 114–121.
<https://doi.org/10.1177/2372732215623554>
- Fives, H., & Gill, M. G. (2014). *International Handbook of Research on Teachers' Beliefs*. Routledge.
<http://ebookcentral.proquest.com/lib/sub/detail.action?docID=1770517>
- Forsberg, E., & Lindberg, V. (2010). *Svensk forskning om bedömning. En kartläggning. [Swedish research on assessment. A review.]* (Vetenskapsrådets rapportserie). The Swedish Research Council (Vetenskapsrådet).
- Fortune, T., & Tedick, D. (n.d.). *Features of Language Objectives for the CBI Classroom*. Retrieved February 6, 2020, from <http://carla.umn.edu/cobalt/modules/curriculum/features.pdf>
- Frankfurt International School. (n.d.). *Assessing ESL students in the subject classroom*. Retrieved April 30, 2020, from <http://esl.fis.edu/teachers/support/eslAssess.htm>
- Fratzke, S. (2017). *Weathering Crisis, Forging Ahead: Swedish Asylum and Integration Policy*. Migration Policy Institute.
<https://www.migrationpolicy.org/sites/default/files/publications/TCM-Asylum-Sweden-FINAL.pdf>
- Fulcher, G. (2016). Testing tasks: Issues in task design and the group oral: *Language Testing*.
<https://doi.org/10.1177/026553229601300103>
- Fulcher, G., & Davidson, F. (2007). *Language Testing and Assessment. An advanced resourcebook*. Routledge.
- Gablasova, D. (2014). Issues in the assessment of bilingually educated students: Expressing subject knowledge through L1 and L2. *The Language Learning Journal*, 42(2), 151–164.
- Gajo, L. (2007). Linguistic knowledge and Subject Knowledge: How does Bilingualism Contribute to Subject Development? *International Journal of Bilingual Education and Bilingualism*, 10(5), 563–581.
- Gallagher, F., & Leahy, A. (2014). The feel good factor: Comparing immersion by design and immersion by default models. *Language, Culture and Curriculum*, 27(1), 58–71.
<https://doi.org/10.1080/07908318.2013.877476>
- García, O. (2008). *Bilingual Education in the 21st Century: A Global Perspective*. Wiley-Blackwell.
- García, O. (2012). *Theorizing translanguaging for educators*. (C. Celic & K. Seltzer, Eds.; Translanguaging: ACUNY-NYSIEB guide for educators, 1-6.).
- García, O., & Wei, L. (2018). *Translanguaging* (C. Chapelle A., Ed.; The Encyclopedia of Applied Linguistics.). John Wiley & Sons, Ltd.
- Gebhard, M. (2019). *Teaching and Researching ELLs' Disciplinary Literacies: Systemic Functional Linguistics in Action in the Context of U.S. School Reform*. Routledge.
<https://doi.org/10.4324/9781315108391>
- Gee, J. P. (2003). Opportunity to Learn: A language-based perspective on assessment. *Assessment in Education: Principles, Policy & Practice*, 10(1), 27–46.
<https://doi.org/10.1080/09695940301696>
- Geijerstam af, Å. (2006). *Att skriva i naturorienterande ämnen i skolan. [Writing in the Natural Sciences in School.]* [Doctoral dissertation]. Uppsala University.
- Geisinger, K. F., Bracken, B. A., Carlson, J. F., Hansen, J.-I. C., Kuncel, N. R., Reise, S. P., & Rodriguez, M. C. (Eds.). (2013). *APA handbook of testing and assessment in psychology, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology*. American Psychological Association. <https://doi.org/10.1037/14047-000>
- Genesee, F., & Lindholm-Leary, K. (2013). Two case studies of content-based language education. *Journal of Immersion and Content-Based Language Education*, 1(1), 3–33.
<https://doi.org/10.1075/jicb.1.1.02gen>

- Gibbons, P. (2002). *Scaffolding language, scaffolding learning: Teaching second language learners in the mainstream classroom*. Heinemann.
- Gibbons, P. (2006). *Stärk språket—Stärk lärandet. [Supporting language, supporting learning.]*. Hallgren & Fallgren.
- Gibbons, P. (2009). *English learners, academic literacy, and thinking: Learning in the challenge zone*. (Heinemann).
- Gibbons, P. (2018). *Bridging Discourses in the ESL Classroom: Students, Teachers and Researchers*. Bloomsbury Academic.
- Gipps, C. (1994). *Beyond Testing. Towards a Theory of Educational Assessment*. Routledge.
- Gipps, C., & Murphy, P. (1994). *A fair test? : Assessment, achievement and equity*.
- Gipps, C., & Stobart, G. (2009). Fairness in Assessment. In C. Wyatt-Smith & J. J. Cumming (Eds.), *Educational Assessment in the 21st Century: Connecting Theory and Practice* (pp. 105–118). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9964-9_6
- Gorard, S. (2007, September 5). *Mixing methods is wrong: An everyday approach to educational justice*. <http://www.leeds.ac.uk/educol/documents/165457.htm>
- Gottlieb, M. (2016). *Assessing English Language Learners: Bridges to Educational Equity (2nd ed.)*. <https://www.ebooks.com/en-us/2509958/assessing-english-language-learners-bridges-to-educational-equity/margo-gottlieb/>
- Government Offices. (2015, July 9). *[Regeringskansliet] Nya nationella skolutvecklingsprogram [Text]*. Regeringskansliet. <https://www.regeringen.se/pressmeddelanden/2015/07/nya-nationella-skolutvecklingsprogram/>
- Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through preservice teacher education. *Teaching and Teacher Education*, 21, 607–621.
- Greene, J. C. (2007). *Mixed methods in social inquiry*. Jossey-Bass. <http://www.loc.gov/catdir/enhancements/fy0739/2007026831-t.html>
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a Conceptual Framework for Mixed-Method Evaluation Designs. *Educational Evaluation and Policy Analysis*, 11(3), 255–274. JSTOR. <https://doi.org/10.2307/1163620>
- Guest, G., & Fleming, P. (2015). Mixed Methods Research. In G. Guest & E. Namey (Eds.), *Public Health Research Methods* (pp. 581–610). SAGE.
- Gunnarsson, K., & Helmersson, K. (2017, March 1). Språkintröskning största gymnasieprogrammet. [Language introduction, the biggest upper secondary program.]. *Sveriges Radio*. <https://sverigesradio.se/sida/artikel.aspx?programid=83&artikel=6641768>
- Hägerfelth, G. (2011). *Språkarbete i alla ämnen*. Liber.
- Hajer, M., & Meestringa, T. (2014). *Spårkinriktad undervisning. [Language oriented instruction.]* (2nd ed.). Hallgren & Fallgren.
- Hall, D. (1995). *Assessing the needs of bilingual pupils: Living in two languages*. David Fulton Publishers.
- Halliday, M. A. K. (1978). Meaning and the construction of reality in early childhood. In H. I. Pick & E. Saltzman (Eds.), *Modes of Perceiving and Processing of Information*. Erlbaum.
- Halliday, M. A. K. (2014). *Halliday's introduction to functional grammar*. Routledge.
- Hamilton, L., & Ravenscroft, J. (2018). *Building Research Design in Education: Theoretically Informed Advanced Methods*. Bloomsbury Publishing.
- Harris, L. R., & Brown, G. T. L. (2009). The complexity of teachers' conceptions of assessment: Tensions between the needs of schools and students. *Assessment in Education: Principles, Policy & Practice*, 16, 365–381.
- Heller, M. (1999). *Linguistic minorities and modernity: A sociolinguistic ethnography*. Longman.

- Hellstén, M., Mickwitz, L., & Scharfenberg, J. (2020). Teacher Education in Sweden: Revisiting the Swedish Model. In K. Pushpanadham & J. Scharfenberg (Eds.), *Teacher Education in Global Era: Perspectives and Practices*. Springer.
- Hennerdal, P., Malmberg, B., & Andersson, E. K. (2018). Competition and School Performance: Swedish School Leavers from 1991–2012. *Scandinavian Journal of Educational Research*, 64(1), 70–86. <https://doi.org/10.1080/00313831.2018.1490814>
- Henriksson, W. (2008). *Statistik. Icke-parametriska metoder. [Statistics. Non-parametric methods.]*. Institutionen för beteendevetenskapliga mätningar, Umeå universitet. <http://umu.diva-portal.org/smash/get/diva2:414804/FULLTEXT03.pdf>
- Hofer, B. (2002). Epistemological world views of teachers: From beliefs to practices. *Issues in Education: Contributions from Educational Psychology*, 8(2), 167–173.
- Höning, I. (2009). *Assessment in CLIL. A case study*. [Magistra der Philosophie]. Universität Wien.
- Hood, P., & Tobutt, K. (2009). *Modern Languages in the Primary School*. SAGE.
- House of Commons Children, Schools and Families Committee. (2008). Testing and Assessment: Third report of Session 2007–08. *Report Together with Formal Minutes, 1*. <http://dera.ioe.ac.uk/7434/2/169ii.pdf>
- Hult, F. M. (2012). English as a Transcultural Language in Swedish Policy and Practice. *TESOL Quarterly*, 46(2), 230–257. JSTOR.
- Hyltenstam, K., & Milani, T. (2012). Flerspråkighetens sociopolitiska och sociokulturella ramar [Multilingualism's sociopolitical and sociocultural framework]. In K. Hyltenstam, M. Axelsson, & I. Lindberg (Eds.), *Flerspråkighet—En forskningsöversikt [Multilingualism—An overview of research]* (pp. 17–152). Vetenskapsrådet.
- Hyltenstam, K., Axelsson, M., Lindberg, I., & Vetenskapsrådet. (2012). *Flerspråkighet: En forskningsöversikt. [Multilingualism: a research review.]*. Vetenskapsrådet.
- IEA. (2019). TIMSS 2019 Science framework. In *TIMSS 2019*. TIMSS & PIRLS Lynch School of Education International Study Center. <http://timss2019.org/wp-content/uploads/frameworks/T19-Assessment-Frameworks-Chapter-2.pdf>
- IES. (2012). *See inside Internationella Engelska Gymnasiet Södermalm / Internationella Engelska Skolan*. <https://engelska.se/news/see-inside-internationella-engelska-gymnasiet-s%C3%B6dermalm>
- Institute for Evaluation of Labour Market and Education Policy. (2014). *[IFAU] Decentralisation, school choice and independent schools: Results and equivalence in the Swedish school system* (No. 25). <https://www.ifau.se/globalassets/pdf/summaries/2015/sum-r-2014-25-decentralisation-school-choice-and-independent-schools.pdf>
- International Bureau of Education. (2016, May 26). *Intended curriculum* [Text]. Bureau International d'éducation. <http://www.ibe.unesco.org/fr/node/12227>
- Jansson, T., Karlstads universitet, & Avdelningen för statsvetenskap. (2011). *Vad kommer på provet?: Gymnasielärares provpraxis i samhällskunskap*. [Licentiate thesis] Fakulteten för samhälls- och livsvetenskaper, Statsvetenskap, Karlstads universitet.
- Johnson, C. C. (2006). Enabling all students to learn science. *The Science Teacher*, 73(2), 32–37.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, 33(7), 14–26. JSTOR.
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a Definition of Mixed Methods Research. *Journal of Mixed Methods Research*, 1(2), 112–133. <https://doi.org/10.1177/1558689806298224>
- Käck, A. (2019). *Digital Competence and Ways of Thinking and Practising in Swedish Teacher Education: Experiences by teachers with a foreign teaching degree*. [Doctoral dissertation] Stockholm University. <http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-171460>

- Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement*. (4th ed). (pp. 17–64). American Council on Education/Praeger.
- Kane, M. (2010). Validity and fairness. *Language Testing*, 27(2), 177–182.
- Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50, 1–73.
- Kaplan, R. B., & Baldauf, R., B. (1997). *Language planning from practice to theory*. Multilingual Matters.
- Kaplan, R. B., & Baldauf, R., B. (2007). Language Policy Spread. Learning from health and social policy models. *Language Problems and Language Planning*, 31(2), 107–129.
- Karlsson, A. (2019). *Det flerspråkiga NO-klassrummet: En studie om translanguaging som läranderesurs i ett NO-klassrum. [The multilingual Natural Science classroom: a study on translanguaging as a resource for learning in a Natural Science classroom.]*. [Docxtoral dissertation] Malmö University.
- Kaya, A. (2017). Betygen blir orättvisa. [The grades become unfair.]. *Skolvärlden*, 1(17).
- Kazamias, A. M. (2009). Comparative Education: Historical Reflections. In R. Cowen & A. M. Kazamias (Eds.), *International Handbook of Comparative Education* (pp. 139–157). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6403-6_10
- Kieffer, M. J., Lesaux, N. K., Rivera, M., & Francis, D. J. (2009). Accommodations for English Language Learners Taking Large-Scale Assessments: A Meta-Analysis on Effectiveness and Validity. *Review of Educational Research*, 79(3), 1168–1201. <https://doi.org/10.3102/0034654309332490>
- Kiely, R. (2009). *CLIL - The question of assessment*. <http://www.developingteachers.com/>
- Kindenberg, B., & Wiksten, M. (2017). *Språkutvecklande NO-undervisning—Strategier och metoder för högstadiet. [A language development perspective in the teaching of the natural sciences—Strategies and methods for secondary education.]*. Natur & Kultur.
- King, N. (2004). Using Templates in the Thematic Analysis of Text. In *Essential Guide to Qualitative Methods in Organizational Research* (pp. 256–270). SAGE Publications Ltd. <https://doi.org/10.4135/9781446280119>
- Kontio, J. (2016). *Auto Mechanics in English. Language Use and Classroom Identity Work* [Doctoral dissertation] Uppsala University. <https://www.avhandlingar.se/om/Janne+Kontio/>
- Köppe, C., & Nijsten, M. (2017, May 15). *Content-Compatible Language/alx*.
- Köppe, C., & Nijsten, M. (2012). *A Pattern Language for Teaching in a Foreign Language—Part 2*. 20.
- Korstjens, I., & Moser, A. (2018). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. *European Journal of General Practice*, 24(1), 120–124. <https://doi.org/10.1080/13814788.2017.1375092>
- Kouns, M. (2010). *INGA IG I KEMI A! En språkdidaktisk studie av en kemilärares undervisningsstrategier i en gymnasieklass med elever med svenska som andraspråk. [No F in chemistry! A language didactic study of a chemistry teacher's teaching strategies in a class of Swedish L2 students in upper secondary school.]* [Licentiate Thesis] Malmö university. <http://muep.mau.se/bitstream/handle/2043/11592/Kouns.pdf;jsessionid=9DAC9AF4AF00335DE2F6B5B2380A4378?sequence=2>
- Kouns, M. (2014). *Beskriv med ord: Fysiklärare utvecklar språkinriktad undervisning på gymnasiet. [Describe with words: Physics teachers develop their language focus in instruction in upper secondary school.]*. [Doctoral dissertation] Malmö university.
- Kroon, S. (2005). *Towards a European Language Monitor. A Contribution to Language Policy Development*. (Lingue, Istituzioni, Territori: Riflessioni teoriche, proposte metodologiche education esperienze di politica linguistica.). Societa Di Linguistica Italiana.

- Labaree, D. F. (2008). *Comments on Bulterman-Bos: The Dysfunctional Pursuit of Relevance in Education Research*. *Educational Researcher*, 37(7), 421–423.
<https://doi.org/10.3102/0013189X08325557>
- Lachat, M. A. (1999). *Standards, Equity and Cultural Diversity*. The Education Alliance, LAB.
- Lambert, W. E. (1955). Measurement of the linguistic dominance of bilinguals. *Journal of Abnormal and Social Psychology*, 50, 197–200.
- Lambert, W. E. (1975). Culture and language as factors in learning and education. In A. Wolfgang (Ed.), *Education of Immigrant Students* (pp. 55–83). Ontario Institute for Studies in Education.
- Lasagabaster, D. (2009). Multilingual educational systems: An added challenge for immigrant students. In J. Miller, A. Kostogriz, & M. Gearon (Eds.), *Culturally and linguistically diverse classrooms* (pp. 18–36). Multilingual Matters.
- Lavigne, A., & Dalal, Y. (2015). Teacher knowledge, beliefs, attitudes. *Classroom Management: An A-to-Z Guide*. <https://doi.org/10.4135/9781483346243.n339>
- Lee, O., Buxton, C., & Shaver, A. (2007). The challenge of altering elementary school teachers' beliefs and practices regarding linguistic and cultural diversity in science instruction. *Journal of Research in Science Teaching*, 44, 1269–1291.
- Leung, C., & Morton, T. (2016). Conclusion: Language Competence, Learning and Pedagogy in CLIL - Deepening and Broadening Integration. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualising Integration in CLIL and Multilingual Education*. Multilingual Matters.
- Li, L. (2013). The complexity of language teachers' beliefs and practice: One EFL teacher's theories. *The Language Learning Journal*, 41(2), 175–191.
<https://doi.org/10.1080/09571736.2013.790132>
- Lim Falk, M. (2008). *Svenska i engelskspråkig skolmiljö: Ämnesrelaterat språkbruk i två gymnasieklasser [Swedish in English speaking school context: Subject language use in two upper secondary classes.]* [Doctoral dissertation] Stockholm University.
<http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-8247>
- Lincoln, Y., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.
- Lindh, C. (2019). *I skrivandets spår: Elever skriver i SO. [In the footsteps of writing: Students write in the Social Sciences.]* [Doctoral dissertation] Malmö University.
<https://doi.org/10.24834/isbn.9789171049964>
- Lindmark, T. (2013). *Samhällskunskapslärares ämneskonceptioner. [Social science teachers' subject conceptions.]* [Doctoral dissertation] Umeå University.
- Lithner, J., Bergqvist, E., & Boesen, J. (2010). Mathematical competences. A research framework. In C. Bersten, W. Jablonka, & T. Wedge (Eds.), *Mathematics and Mathematics education: Cultural and social dimensions Proceedings of MADIF 7, the seventh mathematics education research seminar, Stockholm, January 26-27, 2010*. (pp. 157–167). SMDF.
- Llinares, A., Morton, T., & Whittaker, R. (2012). *The Roles of Language in CLIL*. Cambridge University Press.
- Löfgren, H., Löfgren, R., & Pérez Prieto, H. (2018). Pupils' enactments of a policy for equivalence: Stories about different conditions when preparing for national tests. *European Educational Research Journal*, 17(5), 676–695. <https://doi.org/10.1177/1474904118757238>
- Lyster, R., & Ballinger, S. (2011). Content-based language teaching: Convergent concerns across divergent contexts. *Language Teaching Research*, 15(3), 279–288.
- MAC (Ed.). (2017). *Assessment Literacy Standards A National Imperative*. Michigan Assessment Consortium. https://www.michiganassessmentconsortium.org/wp-content/uploads/MAC_AssessLitStds_2017_9.19.17.pdf

- MacArthur, C. A., Graham, S., & Fitzgerald, J. (2008). *Handbook of Writing Research*. Guilford Press.
- MacSwan, J. (2017). A multilingual perspective on translanguaging. *American Educational Research Journal*, 54(1), 167–201. <https://doi.org/10.3102/0002831216683935>
- Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge*. (pp. 95–132). Kluwer.
- Makoni, S., & Pennycook, A. (2006). *Disinventing and reconstituting languages*. Multilingual Matters.
- Malone, M. E. (2008). Training in Language Assessment. In E. Shohamy & N. H. Hornberger (Eds.), *Language Testing and Assessment* (pp. 225–239). Springer Science+Business Media LLC.
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30, 329–344. <https://doi.org/10.1177/0265532213480129>
- Mariani, L. (1997). Teacher support and teacher challenge in promoting learner autonomy. *Perspectives*, 23(2).
- Martin, J. R. (1989). Technicality and abstraction: Language for the creation of specialised texts. In F. Christie (Ed.), *Writing in Schools* (pp. 36–44). Deakin University Press.
- Martin, J. R. (2002). Writing history: Construing time and value in discourses of the past. In M. J. Schleppegrell & M. C. Colombi (Eds.), *Developing Advanced Literacy in First and Second Languages: Meaning with Power*. Lawrence Erlbaum Associates.
- Martin, J. R., & Rothery, J. (1986). What a functional approach to the writing task can show teachers about “good writing”. In B. Couture (Ed.), *Functional approaches to writing: Research perspectives*. (pp. 241–265). Ablex.
- Massler, U., Stotz, D., & Quessier, C. (2014). Assessment instruments for primary CLIL: the conceptualisation and evaluation of test tasks. *The Language Learning Journal*, 42(2), 137–150.
- Matese, G., Griesdorn, J., & Edelson, D. C. (2002). Teacher beliefs about assessment in inquiry-based science: Research to inform professional development. *American Educational Research Association Conference*.
- Matthews, P. H. (2007). In *The Concise Oxford Dictionary of Linguistics*. Oxford University Press. <https://www.oxfordreference.com/view/10.1093/acref/9780199202720.001.0001/acref-9780199202720-e-61>
- Maxwell, J. A. (2013). *Qualitative research design: An interactive approach*. (3rd ed.). Sage.
- Maxwell, J. A., & Loomis, D. M. (2003). Mixed methods design: An alternative approach. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 241–271). SAGE.
- May, S. (2001). *Language and Minority Rights: Ethnicity, Nationalism and the Politics of Language*. Longman.
- May, S. (Ed.). (2013). *The multilingual turn: Implications for SLA, TESOL, and bilingual education*. Routledge.
- McNamara, T. (2000). *Language Testing*. Oxford University Press.
- McNamara, T., & Roever, C. (2006). *Language Testing: The Social Dimension* (R. Young, Ed.). Blackwell Publishing.
- Menken, K. (2008). *High-stakes Tests as de facto Language Education Policies*. (E. Shohamy & N. H. Hornberger, Eds.; *Language Testing and Assessment*). Springer.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (Vol. 3, pp. 13–103). American Council on Education.
- Met, M. (1997). Integrating language and content: Issues to consider. *Ikastaria*, 9, 27–39.

- Met, M. (1999). *Content-based instruction: Defining terms, making decisions*. (NFLC Reports.). The National Foreign Language Center.
- Moere, A. V. (2012). Paired and Group Oral Assessment. In *The Encyclopedia of Applied Linguistics*. American Cancer Society. <https://doi.org/10.1002/9781405198431.wbeal0894>
- Moere, A. V. (2016). Validity evidence in a university group oral test: *Language Testing*. <https://doi.org/10.1191/0265532206lt3360a>
- Morgan, C. (2006). Appropriate language assessment in content and language integrated learning. *Language Learning Journal*, 33, 59–67.
- Mörtlund, T. (2020). *Arbetet med nyanlända elever i grundskolan. En enkätundersökning om arbetet med nyanlända elever hos kommunala huvudmän för grundskolor.* [Working with newly-arrived students in elementary school. A survey about the work with newly-arrived students in municipal schools.] (2020:2). IFAU, Institutet för arbetsmarknads- och utbildningspolitisk utvärdering. <https://www.ifau.se/globalassets/pdf/se/2020/r-2020-2-arbetet-med-nyanlanda-elever-i-grundskolan.pdf>
- Nagy, W., & Townsend, D. (2012). Words as Tools: Learning Academic Vocabulary as Language Acquisition. *Reading Research Quarterly*, 47(1), 91–108. <https://doi.org/10.1002/RRQ.011>
- National Agency for Education. (n.d.-a). [Skolverket] *Genomföra och bedöma prov i grundskolan.* [Conducting and assessing tests in compulsory school.] [Text]. Retrieved April 30, 2020, from <https://www.skolverket.se/undervisning/grundskolan/nationella-prov-i-grundskolan/genomfora-och-bedoma-prov-i-grundskolan>
- National Agency for Education. (n.d.-b). [Skolverket] *Nationella prov – obligatoriska prov i skolan.* [National tests—Mandatory tests in school.] [Text]. Retrieved June 3, 2020, from <https://www.skolverket.se/for-dig-som-ar.../elev-eller-foralder/betyg-och-nationella-prov/nationella-prov---obligatoriska-prov-i-skolan>
- National Agency for Education. (2005). [Skolverket] *Elever med utländsk bakgrund. En sammanfattande bild.* [Students with migrant background. A summary.].
- National Agency for Education. (2011). [Skolverket] *Barn och elever med utländsk bakgrund. Skolverkets lägesbedömning 2011 Del 1 – Beskrivande data Förskoleverksamhet, skolbarnsomsorg, skola och vuxenutbildning.* [Children and students with migrant background. The National Agency's assessment of the situation 2011. Part 1—Descriptive data Pre school, child care, schools and adult education.].
- National Agency for Education. (2012a). [Skolverket] *Att främja nyanlända elevers kunskapsutveckling.* [Supporting newly-arrived students' knowledge development.] [Text]. <https://www.skolverket.se/publikationsserier/stodmaterial/2012/att-framja-nyanlanda-elevers-kunskapsutveckling>
- National Agency for Education. (2012b). [Skolverket] *Greppa språket! Ämnesdidaktiska perspektiv på flerspråkighet.* [Grasp the language! Disciplinary perspectives on multilingualism.]. Skolverket.
- National Agency for Education. (2015). [Skolverket] *An Assessment of the Situation in the Swedish School System 2015.* 28.
- National Agency for Education. (2016). [Skolverket] *Utbildning för nyanlända elever. Skolverkets allmänna råd med kommentarer.* [Education for newly-arrived students. The National Agency for Education's guidelines with comments.]. <https://www.skolverket.se/system/getfile/getfile>
- National Agency for Education. (2017a). [Skolverket] *Beskrivande data 2017.* [Descriptive data 2017.] (p. 172). Skolverket.
- National Agency for Education. (2017b). [Skolverket] *Elever och skolenheter i grundskolan läsåret 2016/17.* [Students and school units, school year of 2016/17.] (text Dnr 2016:1320). Skolverket. <https://www.skolverket.se/system/getfile/getfile>

- National Agency for Education. (2017c). *[Skolverket] Swedish grades*.
https://www.skolverket.se/download/18.47fb451e167211613ef398/1542791697007/swedishgrades_bilaga.pdf
- National Agency for Education. (2018a). *[Skolverket] Analyser av familjebakgrundens betydelse för skolresultaten och skillnader mellan skolor. En kvantitativ studie av utvecklingen över tid i slutet av grundskolan. [Analyses of the impact of socioeconomic background on school achievements and differences between schools. A quantitative study of the development over time at the end of compulsory school.]* (No. 467). Skolverket.
<https://www.skolverket.se/system/getfile/getfile>
- National Agency for Education. (2018b). *[Skolverket] Betyg och betygssättning. Skolverkets allmänna råd med kommentarer. [Grades and grading. General guidelines from the National Agency for Education.]*. Skolverket. <https://www.skolverket.se/system/getfile/getfile>
- National Agency for Education. (2018c). *[Skolverket] Curriculum for the compulsory school, preschool class and school-age educare 2011 Revised 2018* [Text].
<https://www.skolverket.se/system/getfile/getfile>
- National Agency for Education. (2018d). *[Skolverket] Elever och skolenheter i grundskolan läsåret 2017/2018: [Students and school units, school year of 2017/2018.]* (text Dnr 2018:600).
<https://www.skolverket.se/system/getfile/getfile>
- National Agency for Education. (2018e). *[Skolverket] Greppa flerspråkigheten. [Grasp multilingualism.]*. <https://www.skolverket.se/publikationsserier/forskning-for-skolan/2018/greppa-flersprakigheten>
- National Agency for Education. (2018f). *[Skolverket] Pedagogisk personal i skola och vuxenutbildning läsåret 2017/18. [Staff in schools and adult education the school year of 2017/18.]*. <https://www.skolverket.se/system/getfile/getfile>
- National Agency for Education. (2018g). *[Skolverket] Språk- och kunskapsutvecklande ämnesundervisning för nyanlända elever den första tiden. [Language development in subject content instruction for newly-arrived students during the initial stage.]*
- National Agency for Education. (2019a). *[Skolverket] Elever och skolenheter i grundskolan läsåret 2018/19. [Students and school units in compulsory school, the school year of 2018/19.]* (Dnr: 2018:1562; p. 17).
- National Agency for Education. (2019b). *[Skolverket] Grundskolan—Elevstatistik. [Compulsory school—Student statistics]*.
https://sirir.skolverket.se/reports/rwserverlet?cmdkey=common&geo=1&report=gr_elever&p_sub=1&p_ar=2019&p_lankod=&p_kommunkod=&p_skolkod=&p_hmantyp=&p_hmankod=&p_flik=G
- National Agency for Education. (2019c). *[Skolverket] Gymnasieskolan—Elevstatistik. [Upper secondary school—Student statistics]*.
https://sirir.skolverket.se/reports/rwserverlet?cmdkey=common&geo=1&report=gy_elever&p_ar=2019&p_lankod=&p_kommunkod=&p_skolkod=&p_hmantyp=&p_hmankod=&p_flik=G&p_sub=1
- National Agency for Education. (2019d). *[Skolverket] Pedagogisk personal i skola och vuxenutbildning läsåret 2018/19. [Staff in schools and adult education the school year of 2018/19.]*. Skolverket. <https://www.skolverket.se/system/getfile/getfile>
- National Agency for Education. (2019e). *[Skolverket] Slutbetyg i grundskolan våren 2019. [Final grades in compulsory school spring 2019.]*. <https://www.skolverket.se/system/getfile/getfile>
- National Agency for Education. (2019f). *[Skolverket] TALIS 2018 En studie om lärares och rektorers arbete i grund- och gymnasieskolan. Delrapport 1. [A study about teachers' and principals'*

- work in compulsory school and upper secondary school. Progress report 1.] (No. 481). Skolverket. <https://www.skolverket.se/system/getfile/getfile>
- National Agency for Education. (2020a). [Skolverket] *Gymnasieprogrammen. [The upper secondary school programs.]* [Text]. <https://www.skolverket.se/undervisning/gymnasieskolan/laroplan-program-och-amnen-i-gymnasieskolan/gymnasieprogrammen>
- National Agency for Education. (2020b). [Skolverket] *Sätta betyg i grundskolan. [Grading in compulsory school.]* [Text]. <https://www.skolverket.se/undervisning/grundskolan/betyg-i-grundskolan/satta-betyg-i-grundskolan>
- National Agency for Education. (2020c, March 20). [Skolverket] *Rätt till modersmålsundervisning. [The right to mother tongue instruction.]* [Text]. <https://www.skolverket.se/regler-och-ansvar/ansvar-i-skolfragor/ratt-till-modersmalsundervisning>
- National Agency for Education, (NAE). (2009). [Skolverket] *With another mother tongue—Students in compulsory school and the organisation of teaching and learning.* (No. 321). Skolverket.
- National Agency for Education, (NAE). (2012c). [Skolverket] *Få syn på språket – Ett kommentar material om språk- och kunskapsutveckling i alla skol former, verksamheter och ämnen. [Make language visible—Commentary material regarding language development in all school levels, activities and subjects.]* Skolverket. <https://www.skolverket.se/system/getfile/getfile>
- National Agency for Education, (NAE). (2018h). [Skolverket] *Engelskspråkig undervisning En kartläggning och utvärdering av engelskspråkig undervisning i svensk gymnasieskola. [English-medium instruction. A survey and assessment of English-medium instruction in Swedish upper secondary education.]* (No. 465). Skolverket. <https://www.skolverket.se/system/getfile/getfile>
- National Agency for Education, (NAE). (2018i). [Skolverket] *Nyanlända elever och bedömning i en språk- och kunskapsutvecklande undervisning. [Newly arrived students and assessment in language and knowledge developing instruction].*
- National Agency for Education, (NAE). (2020d, January 17). [Skolverket] *English (engelska)* [Text]. <https://www.skolverket.se/andra-sprak-other-languages/english-engelska>
- Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies*, 19, 317–328.
- Nikula, T. (2017). *CLIL: A European Approach to Bilingual Education: Vol. Second and Foreign Language Education. Encyclopedia of Language and Education* (N. Van Deusen-Scholl & S. May, Eds.; 3rd ed.). Springer, Cham.
- Nikula, T., Dafouz, E., Moore, P., & Smit, U. (Eds.). (2016). *Conceptualising Integration in CLIL and Multilingual Education*. Multilingual Matters.
- Nilsson Folke, J. (2017). *Lived transitions: Experiences of learning and inclusion among newly arrived students*. [Doctoral dissertation] Stockholm University.
- Nilsson, J., & Bunar, N. (2016). Educational Responses to Newly Arrived Students in Sweden: Understanding the Structure and Influence of Post-Migration Ecology. *Scandinavian Journal of Educational Research*, 4, 399. <https://doi.org/10.1080/00313831.2015.1024160>
- Nilsson, Josefin. (2015). *Språk- och kunskapsutvecklande arbetssätt. Lärares lärande: En etnografisk studie om utveckling av språk- och kunskapsutvecklande arbetssätt i undervisning för nyanlända högstadiel elever i ämnet biologi. [Language development teaching. Teachers' learning: An ethnographic study on language development methods in instruction with newly-arrived students in secondary education in biology.]*. <http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-123282>
- Niss, M., & Højgaard, T. (2011). *Competencies and Mathematical Learning: Ideas and inspiration for the development of mathematics teaching and learning in Denmark*. 214.

- Nowak Ottosson, N., & Öhlund, K. (2016). *Kartläggning av nyanlända elevers kunskaper. Specialpedagogens roll i denna process. [Mapping of newly-arrived students' knowledge. The role of the special ed teacher in the process.]* [Examensarbete]. Umeå University. <http://www.diva-portal.se/smash/get/diva2:1079097/FULLTEXT01.pdf>
- Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods*, 16(1), 160940691773384. <https://doi.org/10.1177/1609406917733847>
- Nusche, D., Halász, G., Looney, J., Santiago, P., & Shewbridge, C. (2011). *OECD Reviews of Evaluation and Assessment in Education: Sweden 2011*. OECD. <https://doi.org/10.1787/9789264116610-en>
- Nygård Larsson, P. (2011). *Biologiämnets texter. Text, språk och lärande i en språkligt heterogen gymnasieklass. [Disciplinary texts in biology. Text, language and learning in a linguistically heterogeneous upper secondary class.]* [Doctoral dissertation]. Malmö University.
- Öberg, J. (2019). *Samhällskunskap från styrdokument till undervisning: Tjugo lärare ramar in vad som påverkar deras praktik. [Social sciences, from policy documents to teaching: Twenty teachers' framework that affect their practices.]*
- Odenstad, C., Pettersson, A., Nordgren, K., Stegmann McCallion, M., Bronäs, A., Forskarskolan för lärare i historia och samhällskunskap, Karlstads universitet, & Fakulteten för samhälls- och livsvetenskaper. (2011). *Prov och bedömning i samhällskunskap. En analys av gymnasielärares skriftliga prov. [Tests and assessment in social studies. An analysis of upper secondary teachers written tests.]* <http://urn.kb.se/resolve?urn=urn:nbn:se:kau:diva-6698>
- OECD. (2009a). *Creating effective teaching and learning environments: First results from TALIS*. OECD.
- OECD (Ed.). (2009b). *Creating effective teaching and learning environments: First results from TALIS*. OECD.
- OECD. (2013). Student assessment: Putting the learner at the centre. In OECD, *Synergies for Better Learning* (pp. 139–269). OECD. <https://doi.org/10.1787/9789264190658-7-en>
- OECD. (2014). *Finding the Way: A Discussion of the Swedish Migrant Integration System*. OECD.
- OECD. (2015a). *Helping immigrant students to succeed at school – and beyond*. <https://www.oecd.org/education/Helping-immigrant-students-to-succeed-at-school-and-beyond.pdf>
- OECD. (2015b). *Improving Schools in Sweden: An OECD Perspective*. OECD.
- OECD. (2015c). *PISA-2015-Sweden.pdf*. <https://www.oecd.org/pisa/PISA-2015-Sweden.pdf>
- OECD. (2017). *Education Policy Outlook: Sweden*. <http://www.oecd.org/education/Education-Policy-Outlook-Country-Profile-Sweden.pdf>
- OECD. (2019a). *OECD Economic Surveys SWEDEN Executive Summary March 2019*.
- OECD. (2019b). *PISA 2018 Insights and Interpretations*. <https://www.oecd.org/pisa/PISA%202018%20Insights%20and%20Interpretations%20FINAL%20PDF.pdf>
- Olander, C., & Ingeman, Å. (2011). Towards an inter-language of talking science: Exploring students' argumentation in relation to authentic language. *Journal of Biological Education*, 45(3), 158–164.
- Oliveira, S. (2005). *Language Policy Bernard Spolsky*. https://www.academia.edu/17278430/Language_Policy_Bernard_Spolsky
- Onwuegbuzie, A. J., & Johnson, R. B. (2006). The validity issue in mixed research. *Research in the Schools*, 13(1), 48–63.

- Onwuegbuzie, A. J., & Leech, N. L. (2005). On Becoming a Pragmatic Researcher: The Importance of Combining Quantitative and Qualitative Research Methodologies. *International Journal of Social Research Methodology*, 8(5), 375–387. <https://doi.org/10.1080/13645570500402447>
- Ortega, L. (2018). Ontologies of language, Second Language Acquisition, and world Englishes. *World Englishes*, 37(1), 64–79. <https://doi.org/10.1111/weng.12303>
- Oxford, R. L. (2017). Conditions for Second Language (L2) Learning. In Nelleke Van Deusen-Scholl & S. May (Eds.), *Second and Foreign Language Education* (pp. 27–41). Springer International Publishing. https://doi.org/10.1007/978-3-319-02246-8_4
- Øzker, K. Z. (1998). Vygotskij i tvåspråkighetsforskningen. [Vygotskij in bilingual research.]. In I. Bråten (Ed.), *Vygotskij och pedagogiken*. (pp. 137–163). Studentlitteratur.
- Painter, C. (1999). *Learning through Language in Early Childhood*. Continuum.
- Pajares, M. F. (1992). Teachers' Beliefs and Educational Research: Cleaning Up a Messy Construct. *Review of Educational Research*, 62(3), 307–332. <https://doi.org/10.3102/00346543062003307>
- Panda, A., & Gupta, R. K. (2014). Making academic research more relevant: A few suggestions. *IIMB Management Review*, 26(3), 156–169. <https://doi.org/10.1016/j.iimb.2014.07.008>
- Parkvall, M. (2016). *Sveriges språk i siffror – Vilka språk talas och av hur många. [Sweden's languages in figures—What languages are spoken and by how many.]*. Morfem.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. The National Academies Press. <https://doi.org/10.17226/10019>
- Pérez-Cañado, M. L. (2012). CLIL research in Europe: Past, present, and future. *International Journal of Bilingual Education and Bilingualism*, 15(3), 315–341. <https://doi.org/10.1080/13670050.2011.630064>
- Pitoniak, M., Young, J. W., Martiniello, M., King, T. C., Buteux, A., & Ginsburgh, M. (2009). *Guidelines for the Assessment of English Language Learners*. Educational Testing Service [ETS]. http://www.ets.org/Media/About_ETS/pdf/ELL_Guidelines.pdf (PDF)
- Ponce, O. A., & Pagán-Maldonado, N. (2014). Mixed Methods Research in Education: Capturing the Complexity of the Profession. *International Journal of Educational Excellence*, 1(1), 111–135.
- Popham, W. J. (2004). Why Assessment Illiteracy Is Professional Suicide. *Educational Leadership*, 62(1), 82.
- Popham, W. J. (2006). All About Accountability/Needed: A Dose of Assessment Literacy. *Improving Professional Practice*, 63(6), 84–85.
- Popham, W. J. (2009a). Assessment Literacy for Teachers: Faddish or Fundamental? *Theory Into Practice*, 48, 4–11.
- Popham, W. J. (2009b). Assessment Literacy for Teachers: Faddish or Fundamental? *Theory Into Practice*, 48(1), 4–11. <https://doi.org/10.1080/00405840802577536>
- Porter, A. C. G. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, 31(7), 3–14.
- Porter, A. C. G. (2004). *Curriculum Assessment for Research in Education*. American Educational Research Association.
- Potter, G. (2017). *The Philosophy of Social Science*. (2nd ed.). Routledge.
- Poulson, L., Avramidis, E., Fox, R., Medwell, J., & Wray, D. (2001). The theoretical beliefs of effective teachers of literacy in primary schools: An exploratory study of orientations to reading and writing. *Research Papers in Education*, 16(3), 271–292. <https://doi.org/10.1080/02671520126827>
- Pring, R. (2000a). *Philosophy of Educational Research*. (2nd ed.). Continuum.

- Pring, R. (2000b). The 'False Dualism' of Educational Research. *Journal of Philosophy of Education*, 34(2), 247–260. <https://doi.org/10.1111/1467-9752.00171>
- Reierstam, H. (2015). *Assessing Language or Content? A comparative study of the assessment practices in three Swedish upper secondary CLIL schools* [Licentiate Thesis]. Gothenburg university.
- Reierstam, H. (2018). *Ethics in Educational Research. Dilemmas and ethical considerations in a complex field*. Department of Education, Stockholm University.
- Reierstam, H., & Sylvén, L. K. (2019). Assessment in CLIL. In L. K. Sylvén (Ed.), *Investigating Content and Language Integrated Learning. Insights from Swedish High Schools*. Multilingual Matters.
- Reinhardt, C. S., & Cook, T. D. (1979). Beyond qualitative versus quantitative methods. *Qualitative and Quantitative Methods in Evaluation Research*, 7–32.
- Remesal, A. (2007). Educational reform and primary and secondary teachers' conceptions of assessment: The Spanish instance building upon Black and Wiliam (2005). *Curriculum Journal*, 18, 27–38.
- Richards, J. C., & Schmidt, R. (2002). *Longman Dictionary of Language Teaching and Applied Linguistics*. (3rd ed.). Longman.
- Rubin, M. (2019). *Språkliga redskap—Språklig beredskap: En praktisknära studie om elevers ämnesspråkliga deltagande i luset av inkluderande undervisning. [Linguistic tools—Language readiness: Practical research about students subject language participation in light of inclusive education.]* [Doctoral Thesisrunbvolim].
- Runblom, K. (2018, January 23). Elever nobbar andra främmande språk än engelska. [Students reject other foreign languages but English.]. *Sveriges Radio*. <https://sverigesradio.se/sida/artikel.aspx?programid=83&artikel=6867885>
- SALAR. (2016). *[SKL] Öppna jämförelser – Grundskola. Tema nyanlända elever*. Statens kommuner och landsting. <https://webbutik.skr.se/bilder/artiklar/pdf/7585-392-5.pdf?issuusi=ignore>
- Salö, L. (2010). *Engelska eller svenska? En kartläggning av språksituationen inom högre utbildning och forskning*. Språkrådet.
- Saminathen, M. G., Låftman, S. B., Almquist, Y. B., & Modin, B. (2018). Effective schools, school segregation, and the link with school achievement. *School Effectiveness and School Improvement*, 29(3), 464–484. <https://doi.org/10.1080/09243453.2018.1470988>
- Samuelsson, K. (2019). *Teachers' Work in Times of Restructuring. On Contextual Influences for Collegiality and Professionalism* [Göteborg university]. https://gupea.ub.gu.se/bitstream/2077/59537/2/gupea_2077_59537_2.pdf
- Sandberg, Y. (2018). *Bilingual subject-specific literacies? Teachers' and learners' views and experiences of two school languages in biology, civics, history and mathematics : Case studies from the Swedish upper secondary school*. [Licentiate thesis] Stockholm University. <http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-159542>
- Sandelowski, M. (2000). Combining Qualitative and Quantitative Sampling, Data Collection, and Analysis Techniques in Mixed-Method Studies. *Research in Nursing & Health*, 23(3), 246–255. [https://doi.org/10.1002/1098-240X\(200006\)23:3<246::AID-NUR9>3.0.CO;2-H](https://doi.org/10.1002/1098-240X(200006)23:3<246::AID-NUR9>3.0.CO;2-H)
- Schissel, J. L., Leung, C., & Chalhoub-Deville, M. (2019). The Construct of Multilingualism in Language Testing. *Language Assessment Quarterly*, 16(4–5), 373–378. <https://doi.org/10.1080/15434303.2019.1680679>
- Schleppegrell, M. J. (2001). Linguistic features of the language of schooling. *Linguistics and Education*, 12(4), 431–459.
- Schleppegrell, M. J. (2004). *The Language of Schooling. A Functional Linguistics Perspective*. Lawrence Erlbaum Associates.

- Schlepppegrell, M. J. (2005). The challenges of academic language in school subjects. In I. Lindberg & K. Sandwall (Eds.), *Språket och kunskapen—Att lära på sitt andraspråk i skola och högskola. Rapport från nordisk konferens den 7-8 oktober i Göteborg. [Language and knowledge—To learn in a second language at school and university. Report from Nordic conference 7-8 October in Gothenburg.]* (Institutet för svenska som andraspråk, Göteborgs universitet.).
- Schlepppegrell, M. J. (2011). Teaching Academic Language in L2 Secondary Settings. *Annual Review of Applied Linguistics*, 31, 3–18. <https://doi.org/10.1017/S0267190511000067>
- Schlepppegrell, M. J. (2020). The knowledge base for language teaching: What is the English to be taught as content? *Language Teaching Research*, 24(1), 17–27. <https://doi.org/10.1177/1362168818777519>
- Schoonenboom, J., & Johnson, R. B. (2017). How to Construct a Mixed Methods Research Design. *Kölner Zeitschrift Für Soziologie Und Sozialpsychologie*, 2(69), 107–131. <https://doi.org/10.1007/s11577-017-0454-1>
- Scott, D. (2007). Resolving the quantitative–qualitative dilemma: A critical realist approach. *International Journal of Research & Method in Education*, 30(1), 3–17. <https://doi.org/10.1080/17437270701207694>
- Scott, D., & Beadle, S. (2014). *Improving the effectiveness of language learning: CLIL and computer assisted language learning*. 44.
- Scott, S., Webber, C. F., Lupart, J. L., Aitken, N., & Scott, D. E. (2014). Fair and equitable assessment practices for all students. *Assessment in Education: Principles, Policy & Practice*, 21(1), 52–70. <https://doi.org/10.1080/0969594X.2013.776943>
- Scully, D. (2017). Constructing Multiple-Choice Items to Measure Higher-Order Thinking. *Practical Assessment Research and Evaluation*, 22(4).
- SFS 2010:800. (2010). *Skollagen 2010:800 [The Swedish Education Act, SFS 2010:800]*. Utbildningsdepartementet [Ministry of Education and Research]. https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/skollag-2010800_sfs-2010-800
- SFS 2011:185. (2011). *Skolförordning 2011:185 [The Ordinance for Compulsory School 2011:185]*. Utbildningsdepartementet [Ministry of Education and Research]. https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/skolforordning-2011185_sfs-2011-185
- Sharif, H. (2017). *"Här i Sverige måste man gå i skolan för att få respekt" Nyanlända ungdomar i den svenska gymnasieskolans introduktionsutbildning. ["Here in Sweden you have to go to school to get respect" Newly arrived students in the Swedish upper secondary introductory education.]* [Doctoral dissertation] Uppsala University. <https://uu.diva-portal.org/smash/get/diva2:1088592/FULLTEXT01.pdf>
- Shephard, L. A. (2000). *The role of classroom assessment in teaching and learning*. (No. 517). National Center for Research on Evaluation, Standards, and Student Testing.
- Shohamy, E. (2003). Implications of language education policies for language study in schools and universities. *Modern Language Journal*, 87(ii), 278–296.
- Shohamy, E. (2006). *Language Policy: Hidden Agendas and New Approaches*. Routledge. <http://ebookcentral.proquest.com/lib/sub/detail.action?docID=258983>
- Shohamy, E. (2008). *Language Testing and Assessment*. (E. Shohamy & N. H. Hornberger, Eds.; Vol. 7). Springer Science+Business Media LLC.
- Shohamy, E. (2011). Assessing Multilingual Competencies: Adopting Construct Valid Assessment Policies. *The Modern Language Journal*, 95(3), 418–429. <https://doi.org/10.1111/j.1540-4781.2011.01210.x>

- Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>
- Siegel, M. A. (2007). Striving for equitable classroom assessments for linguistic minorities: Strategies for and effects of revising life science items. *Journal of Research in Science Teaching*, 44(6), 864–881.
- Siegel, M. A., Wissehr, C., & Halverson, K. (2008). Sounds like success: A framework for equitable assessment. *The Science Teacher*, 75(3), 43–46.
- Sireci, S. G., & Faulkner-Bond, M. (2015). Promoting Validity in the Assessment of English Learners. *Review of Research in Education*, 39, 215–252.
- Sireci, S. G., Li, S., & Scarpatti, S. (2003). *The effects of tests accommodations on test performance: A review of the literature. Commissioned paper by the National Academy of Sciences/National Research Council's Board on Testing and Assessment*. School of Education, University of Massachusetts Amherst. Center for Educational Assessment Research Report no. 485.
- Sireci, S. G., Scarpatti, S. E., & Li, S. (2016). Test Accommodations for Students With Disabilities: An Analysis of the Interaction Hypothesis. *Review of Educational Research*. <https://doi.org/10.3102/00346543075004457>
- Sireci, S. G., & Sukin, T. (2013). Test validity. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology*. (pp. 61–84, Chapter xxix, 711 Pages). American Psychological Association (Washington, DC, US). <http://search.proquest.com/docview/1547568793/abstract/C1DAA1CC64264520PQ/1>
- Slaney, K. L. (2012). Laying the cornerstone of construct validity theory: Herbert Feigl's influence on early specifications: *Theory & Psychology*. <https://doi.org/10.1177/0959354311400659>
- Smeyers, P. (2006). The Relevance of Irrelevant Research; The Irrelevance of Relevant Research. In P. Smeyers & M. Depaepe (Eds.), *Educational Research: Why 'What Works' Doesn't Work* (pp. 95–108). Springer Netherlands. https://doi.org/10.1007/978-1-4020-5308-5_6
- Smeyers, P., & Depaepe, M. (2010). *Educational Research – the Ethics and Aesthetics of statistics*. Dordrecht: Springer Netherlands. Springer Netherlands.
- Snow, M. A., Met, M., & Genesee, F. (1989). A Conceptual Framework for the Integration of Language and Content in Second/Foreign Language Instruction. *TESOL Quarterly*, 23(2), 201–217. <https://doi.org/10.2307/3587333>
- SOU 2000:92. (2000). *Advantage Sweden—En kraftsamling inför ökad rekrytering av utländska studenter till Sverige. [Joint forces for increased enrollment of international students in Sweden.]*. Utbildningsdepartementet [Ministry of Education and Research]. <https://www.regeringen.se/48d068/contentassets/4d0a55e1bde64e3a9e34480bd035adbc/sou-200092-advantage-sweden---en-kraftsamling-for-okad-rekrytering-av-utlandska-studenter-till-sverige>
- SOU 2008:26. (2008). *Värna språken: Förslag till språklag: betänkande. [Defend the languages: language act proposition: report*. Fritze.
- SOU 2017:54. (2017). *Fler nyanlända elever ska uppnå behörighet till gymnasiet. [Make more newly-arrived students eligible for upper secondary education.]*. Wolters Kluwer.
- SOU 2018:78. (2018). *Internationalisation of Swedish Higher Education Institutions*. Utbildningsdepartementet [Ministry of Education and Research]. <https://www.regeringen.se/4ab6ff/contentassets/2cf334f2792c42d4b8c414076eb59204/summary-in-english-internationalisation-of-swedish-higher-education-institutions>
- Spolsky, B. (1977). *Language Testing: An Art or Science*. 7(28).
- Spolsky, B. (1995). *Measured Word*. Oxford University Press.

- Spolsky, B. (2004). *Language policy*. Cambridge University Press.
- Spolsky, B., & Shohamy, E. (2000). *Language Practice, Language Ideology, and Language Policy*. (R. D. Lambert & E. Shohamy, Eds.; *Language Policy and Pedagogy: Essays in Honor of A. Ronald Walton*.) John Benjamins.
- Statistics Sweden. (2017). *[SCB] Lärare utanför yrket*. Statistiska centralbyrån.
- Statistics Sweden. (2019). *[SCB] Befolkningsstatistik*. Statistiska centralbyrån.
<http://www.scb.se/hitta-statistik/statistik-efter-amne/befolkning/befolkningens-sammansattning/befolkningsstatistik/>
- Stearns, P. N., Seixas, P. C., & Wineburg, S. S. (2000). *Knowing, teaching, and learning history: National and international perspectives*. New York University Press.
- Stobart, G. (2005). Fairness in multicultural assessment systems. *Assessment in Education: Principles, Policy & Practice*, 12(3), 275–287. <https://doi.org/10.1080/09695940500337249>
- Strömberg Jämsvi, S. (2019). *Unpacking dominant discourses in higher education language policy. A critical study of language policy in Swedish higher education* [Doctoral dissertation] Gothenburg University.
https://gupea.ub.gu.se/bitstream/2077/58606/2/gupea_2077_58606_2.pdf
- Sugrue, C. (1997). Student teachers' lay theories and teaching identities: Their implications for professional development. *European Journal of Teacher Education*, 20, 213–225.
- Sundqvist, P., & Sylvén, L. K. (2016). *Extramural English in Teaching and Learning*. Palgrave Macmillan.
- Suter, L. E., Smith, E., & Denman, B. D. (2019). *The SAGE Handbook of Comparative Studies in Education*. <https://doi.org/10.4135/9781526470379>
- Svartvik, J. (1999). *Engelska: Öspråk, världsspråk, trendspråk. [English: Island language, world language, trend language]*. P.A. Norstedt & Söner AB.
- Svensson Källberg, P. (2018). *Immigrant students' opportunities to learn mathematics In(ex)clusion in mathematics education* [Doctoral dissertation] Stockholm University. <http://www.diva-portal.org/smash/get/diva2:1177002/FULLTEXT01.pdf>
- Swedish Council for Higher Education. (2019, April 8). *[Universitets och högskolerådet, UHR] Språkrkrav för utbildningar som ges på engelska. [Language requirements for education given in English.]* <https://antagning.se/sv/Hur-du-anmaler-dig-och-haller-koll/magister-och-master/sprakkrav-for-utbildningar-som-ges-pa-engelska/>
- Swedish National Agency for School Improvement. (2007). *[Myndigheten för skolutveckling.] Nationell strategi för utbildning av nyanlända barn och ungdomar. Arbetsmaterial 070404. [National strategy for the education of newly-arrived children and youth. Work-in-progress 070404.]*
www.skolutveckling.se/mangfald_likvardighet/nyanlanda/nationell_strategi/skolintroduktion.
- Swedish National Agency for School Improvement. (2018). *[Skolforskningsinstitutet] Språk- och kunskapsutvecklande undervisning i det flerspråkiga klassrummet – med fokus naturvetenskap [Language development in instruction in the multilingual classroom - focus on the natural sciences]*.
- Swedish Research Council, [Vetenskapsrådet]. (2017). *Good Research Practice*.
<https://www.vr.se/english/analysis/reports/our-reports/2017-08-31-good-research-practice.html>
- Swedish School Inspectorate. (2009). *[Skolinspektionen] Utbildning för nyanlända elever. [Education for newly-arrived students.]* (No. 3; p. 34).
- Swedish School Inspectorate. (2017). *[Skolinspektionen] Bedömningsprocessernas betydelse för likvärdigheten. Ombedömning av nationella prov 2016. Redovisning av regeringsuppdrag Dnr U2014/7535/GV [The impact of the assessment process for equity. Reassessment of*

- national tests 2016. Presentation of governmental assignment*] [Dnr U2014/7535/GV]. Skolinspektionen.
- Swedish School Inspectorate. (2018). *[Skolinspektionen] Rättvis och likvärdig betygssättning i grundskolan? Granskning av rektors och huvudmäns arbete. [Fair and equitable grading in compulsory school? Inspection of principals' and school organizers' work.]*. https://www.skolinspektionen.se/globalassets/publikationssok/granskningsrapporter/kvalitetssgranskningar/2018/betygssattning/rattvis-och-likvardig-betygssattning-i-grundskolan_rapport_skolinspektionen_2018.pdf
- Sylvén, L. K. (2005). Extramural exposure to English among Swedish school students. In *Approaches to teaching and learning in linguistic research: Papers from the ASLA symposium in Växjö, 10-11 November 2005* (pp.252-260). ASLA, svenska föreningen för tillämpad språkvetenskap.
- Sylvén, L. K. (Ed.). (2019). *Investigating Content and Language Integrated Learning. Insights from Swedish High Schools*. Multilingual Matters.
- Sylvén, L. K., & Ohlander, S. (2014). The CLISS Project: Receptive Vocabulary in CLIL versus non-CLIL Groups. *Moderna Språk*, 108(2), 80–114.
- Symonds, J., E., & Gorard, S. (2008). *The Death of Mixed Methods: Research Labels and their Casualties*. The British Educational Research Association Annual Conference, Heriot Watt University, Edinburgh, September 3-6, Edinburgh, Scotland. <https://www.leeds.ac.uk/educol/documents/174130.pdf>
- Tamimy, M. (2015). Consistencies and Inconsistencies Between Teachers' Beliefs and Practices. *The Qualitative Report* 2015, 20(8), 1234–1259.
- Tarone, E. (1998). Research on Interlanguage Variation. In L. F. Bachman & A. Cohen D. (Eds.), *Interfaces Between Second Language Acquisition and Language Testing Research*. (pp. 71–89). Cambridge University Press.
- Teleman, U. (2003). *Tradis och funkis: Svensk språkvård och språkpolitik efter 1800. [Traditionalism and functionalism: Swedish language cultivation and politics after 1800]*. Norstedts Ordbok.
- The National Encyclopedia (NE). (n.d.). *Modersmålsundervisning—Uppslagsverk—NE.se*. Retrieved February 20, 2020, from <https://www.ne.se/uppslagsverk/encyklopedi/l%C3%A5ng/modersm%C3%A5lsundervisning>
- Theens, F. (2019). *Does language matter?: Sources of inequivalence and demand of reading ability of mathematics tasks in different languages*. Umeå University.
- Thijs, A., & Akker, J. V. D. (Eds.). (2009). *Curriculum in development*. SLO– Netherlands Institute for Curriculum Development.
- Thomas, W., & Collier, V. (2002). *A National Study of School Effectiveness for Language Minority Students' Long-Term Academic Achievement*. <https://escholarship.org/uc/item/65j213pt>
- Thorne, S. L., & Lantolf, J. P. (2006). A Linguistics of Communicative Activity. In S. Makoni & A. Pennycook (Eds.), *(Dis)inventing and reconstituting languages*. Multilingual Matters.
- Tierney, R. D., Simon, M., & Charland, J. (2011). Being Fair: Teachers' Interpretations of Principles for Standards-Based Grading. *The Educational Forum*, 75(3), 210–227. <https://doi.org/10.1080/00131725.2011.577669>
- Tobin, G. A., & Begley, C. M. (2004). Methodological rigour within a qualitative framework. *Journal of Advanced Nursing*, 48(4), 388–396. <https://doi.org/10.1111/j.1365-2648.2004.03207.x>
- Toth, J. (2018). *English-medium instruction for young learners in Sweden A longitudinal case study of a primary school class in a bilingual EnglishSwedish school* [Doctoral dissertation] Stockholm University. <https://su.diva-portal.org/smash/get/diva2:1205724/FULLTEXT01.pdf>
- Toulmin, S. E. (2003). *The Uses of Argument*. Cambridge University Press.

- Uddling, J. (2019). *Textsamtalens möjligheter och begränsningar i språkligt heterogena fysikklassrum. [The possibilities and limitations of text talks in linguistically diverse classrooms.]* [Doctoral dissertation] Stockholm university.
- Umansky, I. M., Valentino, R. A., & Reardon, S. F. (2016). The Promise of Two-Language Education—Educational Leadership. *Helping ELLs Excel*, 73(5), 10–17.
- Ünsal, Z. (2017). *Bilingual Students' Learning in Science Language, Gestures and Physical Artefact* [Doctoral dissertation] Stockholm University.
- van Lier, L. (1989). Reeling, Writhing, Drawling, Stretching, and Fainting in Coils: Oral Proficiency Interviews as Conversation. *TESOL Quarterly*, 23(3), 489–508. JSTOR. <https://doi.org/10.2307/3586922>
- Varaki, B. S., Floden, R. E., & Kalatehjafarabadi, T. J. (2015). Para-quantitative Methodology: Reclaiming experimentalism in educational research. *Open Review of Educational Research*, 2(1), 26–41. <https://doi.org/10.1080/23265507.2014.986189>
- Verma, S. (2014). *Development of Life Skills and Professional Practice (WBSCTE)*. Vikas Publishing House, PVT LTD.
- Webb, N. L. (1997). *Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education*. 47.
- Webb, N. L. (2002). Assessment Literacy in a Standards-Based Education Setting. *WCER Working Paper No. 2002-4*, 21.
- Wernesjö, U. (2014). *Conditional Belonging Listening to Unaccompanied Young Refugees' Voices* [Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences 93, Uppsala University]. <http://www.diva-portal.org/smash/get/diva2:689776/FULLTEXT01.pdf>
- Wewer, T. (2014). *Assessment of Young Learners' English Proficiency in Bilingual Content Instruction CLIL*. [Doctoral dissertation] University of Turku, Faculty of Education/Department of Teacher Education.
- White, E. (Ed.). (2018). *Handbook of Research on Assessment Literacy and Teacher-Made Testing in the Language Classroom*. IGI Global.
- Wikström, C. (2014). *Konsten att göra bra prov. [The art of developing good tests.]*. Natur & Kultur.
- Williams, C. (1994). *Arfarniadoddulliaudysguacaddysguynghyddestunaddysguwchraddddwyieithog [An evaluation of teaching and learning methods in the context of bilingual secondary education] [Unpublished doctoral thesis]* University of Bangor, Wales.
- Wilson, K., & Devereux, L. (2014). Scaffolding theory: High challenge, high support in academic language and learning (ALL) contexts. *Journal of Academic Language & Learning*, 8(3), A91–A100.
- Wittgenstein, L. (1921). *Tractatus Logico-Philosophicus*.
- Wolff, D. (2009). Content and language integrated learning. In K. Knapp & B. Seidlhofer (Eds.), *Handbook of Foreign Language Communication and Learning*. (pp. 545–572). Mouton de Gruyter.
- Xu, Y., & Brown, G. (2016). Teacher Assessment Literacy in Practice: A Reconceptualization. *Teaching and Teacher Education*, 58, 149–162. <https://doi.org/10.1016/j.tate.2016.05.010>
- Young, J. W. (2009). A Framework for Test Validity Research on Content Assessments Taken by English Language Learners. *Educational Assessment*, 1532–6977.
- Young, J. W., King, T. C., Hauck, M. C., Ginsburgh, M., Kotloff, L., Cabrera, J., & Cavalie, C. (2014). Improving Content Assessment for English Language Learners: Studies of the Linguistic Modification of Test Items. *ETS Research Report Series*, 2014(2), 1–79. <https://doi.org/10.1002/ets2.12023>

- Young, M. (2013). Overcoming the Crisis in Curriculum Theory: A Knowledge-Based Approach. *Journal of Curriculum Studies*, 45(2), 101–118.
- Young, R. (1995). Conversational Styles in Language Proficiency Interviews. *Language Learning*, 45(1), 3–42. <https://doi.org/10.1111/j.1467-1770.1995.tb00961.x>
- Yoxsimer Paulsrud, B. (2014). *English-medium instruction in Sweden: Perspectives and practices in two upper secondary schools*. [Doctoral dissertation] Stockholm university. <http://urn.kb.se/resolve?urn=urn:nbn:se:du-14437>

Appendices

Appendix A: Missivbrev på svenska

(original)

Stockholm, [aktuell månad] 2018

Hej,

Jag kontaktar dig som undervisar något NO- eller SO-ämne och har nyanlända elever i din undervisning.

Jag heter Helena och är doktorand vid Stockholms universitet, institutionen för pedagogik och didaktik. Efter min licentiatavhandling om lärares bedömningspraktik på engelskspråkiga gymnasieskolor, fortsätter jag nu i min doktorsavhandling att fokusera på hur det ser ut i skolor med nyanlända elever.

Med detta mail kommer en förfrågan till dig som lärare att besvara en enkät och därmed ingå i och bidra till min forskningsstudie. Studien handlar om hur det är att bedöma ämneskunskap bland nyanlända elever. Enkäten uppskattas ta ungefär 15-20 minuter och kommer behandlas helt anonymt. Jag hoppas därför att du kan tänka dig att bidra genom att svara på enkäten, det är bara att klicka på länken nedan:

<https://survey.su.se/Survey/22982>

När det gäller bedömning av ämneskunskap bland elever med ett annat modersmål, finns det väldigt lite forskning, både i Sverige och internationellt. En studie som denna är därför oerhört viktig och efterfrågad. Genom att svara på enkäten bidrar du till att vidga förståelsen av hur situationen ser ut i Sverige idag, inte minst med tanke på en ökande språklig mångfald och varierande språkförmåga bland elever, särskilt med tanke på nyanlända.

Jag uppskattar verkligen att du lägger tid på detta och vore tacksam om du kan besvara enkäten så snart som möjligt då dina svar läggs samman med enkätsvar från lärare som svarade i slutet av läsåret 2017/18. Vissa frågor syftar tillbaka på dina erfarenheter från förra läsåret (2017/18), medan andra frågor syftar på din erfarenhet mer generellt.

Med tacksamma hälsningar,
Helena

Appendix B: Letter in English

(translated from Swedish original)

Stockholm, [current month] 2018

Hello,

I am contacting you because you teach a natural or social science subject and have newly-arrived students in your classes.

My name is Helena, and I am a PhD student at Stockholm university in the Department of Education. Following my licentiate thesis on teachers' assessment practices in English-medium upper secondary schools, my focus for my doctoral thesis is on schools with newly-arrived students.

With this email I therefore ask you as a teacher to answer a survey and thereby be a part of and contribute to my investigation. The study concerns assessment of subject content among newly-arrived students. The survey takes about 15–20 minutes and will be treated anonymously. My hope is that you will be willing to contribute by answering the survey; all you need to do is click on the link below:

<https://survey.su.se/Survey/22982>

When it comes to assessment of subject content among students with a different mother tongue, there is very little research to date in either Sweden or internationally. Therefore, a study such as the present one is very important and in-demand. By answering the survey, you contribute to a better understanding of the situation in Sweden today, especially as regards growing multilingualism and varied language proficiency among students, particularly the newly-arrived.

I truly appreciate you taking the time and I would be grateful if you could answer the survey as soon as possible, since your answers will be combined with answers from teachers who responded at the end of the 2017/18 school year. Certain questions refer to your experiences from last year (2017/18), while other questions are more generic.

Kind regards,
Helena

Appendix C: Enkät på svenska

(original)



Stockholms
universitet

Del 1. Bakgrundsfrågor

Hej,

Syftet med denna enkät är att få insyn i hur ämneslärares vardag ser ut vid bedömning av ämneskunskap bland nyanlända elever. Svaren kommer att användas i en avhandling som undersöker lärares bedömningspraktik bland nyanlända och andraspråkinlärare.

Enkäten vänder sig till dig som undervisar i något SO- eller NO-ämne i grundskolans senare årskurser, på språkintruktionsprogram eller på gymnasiet. Alla svar kommer att behandlas anonymt.

I slutet av enkäten finns möjlighet att lämna dina kontaktuppgifter ifall du kan tänka dig att bli intervjuad, ungefär 20 minuter online, per telefon eller "live". Detta för att ge mig djupare insikt i hur du ser på bedömning i din undervisning.

Stort tack för att du tar dig tid att besvara enkäten. Dina svar är viktiga!

Helena Reierstam
helena.reierstam@edu.su.se

1. Vilken skolförundervisar du i? Välj det/de alternativ som passar.

☐ Grundskola åk 7-9

☐ Introduktionsprogram

☐ Gymnasieskola

☐ Annat, specificera:

2. Vilken kommun arbetar du i?

3. Är du man eller kvinna?

☐ Man

☐ Kvinna

☐ Annat/Vill inte ange

4. Hur många år har du arbetat som lärare?

5. Ditt födelseår:

6. Har du svensk lärarlegitimation?

☐ Ja

☐ Nej

Kommentar

7. Vilket år tog du din lärarexamen? Om lärarexamen saknas, skriv saknas.

8. Vilken utbildning har du? Välj det/de alternativ som passar.

☐ Lärarutbildning grundskola

☐ Lärarutbildning gymnasieskola

☐ Annan utbildning

Om annan, specificera:

9. Har du annan yrkeserfarenhet?

☐ Ja

☐ Nej

Om ja, ange vilken:

10. Vilket/vilka ämnen undervisade du i föregående läsår (2017/18) och vilka ämnen har du behörighet att undervisa i? Kryssa i båda kolumnerna utifrån din bakgrund. Ett eller flera alternativ.

	Undervisar i:	Behörig i:
Räddning	<input type="checkbox"/>	<input type="checkbox"/>
Fysik	<input type="checkbox"/>	<input type="checkbox"/>
Kemi	<input type="checkbox"/>	<input type="checkbox"/>
Teknik	<input type="checkbox"/>	<input type="checkbox"/>
Naturkunskap	<input type="checkbox"/>	<input type="checkbox"/>
Matematik	<input type="checkbox"/>	<input type="checkbox"/>
Historia	<input type="checkbox"/>	<input type="checkbox"/>
Geografi	<input type="checkbox"/>	<input type="checkbox"/>
Religion	<input type="checkbox"/>	<input type="checkbox"/>
Samhällskunskap	<input type="checkbox"/>	<input type="checkbox"/>
Svenska	<input type="checkbox"/>	<input type="checkbox"/>
Svenska som andraspråk	<input type="checkbox"/>	<input type="checkbox"/>
Sfi	<input type="checkbox"/>	<input type="checkbox"/>
Engelska	<input type="checkbox"/>	<input type="checkbox"/>
Annat/andra	<input type="checkbox"/>	<input type="checkbox"/>

Om annat/andra, ange vilket/vilka:

Del 2. Undervisningskontext

I denna enkät används begreppen **nyanlända elever** och **andraspråksinlärare**. *Nyanlända* definieras som elever som har gått i svensk skola kortare tid än fyra år, i överensstämmelse med skollagens definition. *Andraspråksinlärare* syftar på elever med ett annat modersmål än svenska som behöver språklig stöttning. Nyanlända ingår i båda kategorierna medan andraspråksinlärare inte behöver vara nyanlända.

Frågorna handlar mest om nyanlända elever, men ibland efterfrågas erfarenhet av andraspråksinlärare mer generellt.

11. Hur länge har du haft nyanlända elever i din undervisning? Se definition av nyanlända ovan.

- ☐ 1 år
☐ 2-3 år
☐ 4-5 år
☐ 6 år eller mer

Kommentar

12. Innan du tog emot nyanlända, har du haft andraspråksinlärare med ett annat modersmål än svenska i din undervisning? Se definition ovan.

- ☐ Aldrig
☐ Ja, någon enstaka gång
☐ Ja, då och då
☐ Ja, hela tiden
☐ Vet ej

Kommentar

13. Ungefär hur många olika språk fanns totalt representerade bland nyanlända och andraspråksinlärare i dina samtliga klasser föregående läsår (2017/18)? Gör en sammanfattande skattning. Välj ett alternativ.

- ☐ 1-2
☐ 3-5
☐ 6-9
☐ 10 eller fler

Kommentar

14. Ange genomsnittlig andel elever som var andraspråksinlärare i dina undervisningsgrupper under föregående läsår (2017/18) och hade ett annat modersmål än svenska. Kryssa för den/de skolform(er) som är relevanta för dig.

	Ett fåtal	En fjärdedel	Hälften	Tre fjärdedelar	De flesta
Grundskola	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Språkintrödnktion	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Gymnasieskola	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Kommentar

Del 3. Nyanlända och undervisning

Tänk på en av dina mest typiska undervisningsgrupper med **större andel nyanlända** under föregående läsår (2017/18) och svara på följande frågor.

Beskriv först vilket ämne och årskurs/skolform det gäller, t ex NO, åk 9 eller biologi, språkintrödnktion/SPRINT. Hur många elever som gick i gruppen, t ex 25, och hur många elever i gruppen som var nyanlända, t ex 8.

15. Ämne:

16. Årskurs och/eller skolform:

17. Totalt antal elever i gruppen:

18. Antal nyanlända i gruppen:

19. Hur ofta gav du samtliga elever i gruppen följande typ av uppgift? Välj det som stämmer bäst på en skala från nästan aldrig till flera gånger under kursen.

	Nästan aldrig	Någon gång under kursen	Flera gånger under kursen
Lösa en kort uppgift enskilt under en lektion	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lösa en kort uppgift tillsammans i mindre grupp under en lektion	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Göra en längre skrivuppgift enskilt under några veckor	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Göra en längre skrivuppgift i grupp under några veckor	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Göra en förberedd muntlig redovisning enskilt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Göra en förberedd muntlig redovisning i grupp	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Kommentar

20. Hur svårt eller lätt anser du att följande är för dina nyanlända elever? Kryssa på en skala från svårt till lätt.

	Svårt	Ganska svårt	Varken lätt eller svårt	Ganska lätt	Lätt
Muntliga ämnesdiskussioner i klassrummet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Muntliga redovisningar efter förberedelse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Korta skriftliga uppgifter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Längre skrivuppgifter, t ex uppsatser och rapporter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Kommentar

21. Hur svårt eller lätt anser du att följande är för dina nyanlända elever när de ska uttrycka sig? Kryssa på en skala från svårt till lätt.

	Svårt	Ganska svårt	Varken lätt eller svårt	Ganska lätt	Lätt
Ämnesspecifika ord och begrepp	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ämnesrelaterade förmågor, t ex kunna argumentera, dra slutsatser	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ämnesspecifik muntlig genre	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ämnesspecifik skriftlig genre	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Allmänt ämnesövergripande skolspråk	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vardagsspråk	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Kommentar

22. Hur ofta anser du att dina nyanlända elever har svårt att visa sitt ämneskunnande i följande aktiviteter? Kryssa på en skala från alltid till aldrig.

	Alltid	Ofta	Ibland	Sällan	Aldrig
Argumentera	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Analysera	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Beskriva/redogöra för	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Diskutera	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Dra slutsatser	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Formulera frågeställningar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Förklara	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Föra resonemang	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Göra jämförelser	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Identifiera/känna igen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Motivera	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Namnge	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Utvärdera/värdera	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Visa på samband	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Återge	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Kommentar

Sedan den 15 april 2016 finns Skolverkets kartläggningssmaterial steg 1 och 2 för att ge en bild av nyanlända elevers utbildningsbakgrund, språkanvändning och intressen. Steg 3 är tänkt som ett stöd för ämneslärare i undervisningen med nyanlända i olika ämnen.

23. Ange det som är relevant i ditt fall avseende kartläggning.

	Ja	Delvis	Nej	Vet ej
Har du tagit del av resultat från kartläggning steg 1 och 2?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Har du genomfört kartläggning steg 3 i ditt ämne?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Har du tagit del av kartläggning steg 3 som någon annan har genomfört?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Om du har tagit del av resultat från kartläggning, på vilket sätt har det hjälpt dig?

Del 4. Bedömning

24. Hur ofta använder du följande då du bedömer elevers kunskap i ditt ämne/dina ämnen? Nyanlända och/eller andra elever. Kryssa på en skala från alltid till aldrig.

	Alltid	Ofta	Ibland	Sällan	Aldrig
Egna prov och uppgifter som jag har konstruerat	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Andra (kollegors) prov och uppgifter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Standardiserade prov, t ex gamla nationella prov	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Skolverkets aktuella nationella prov och/eller bedömningsstöd	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Prov i digital form	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Bedömning av muntliga uppgifter/presentationer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Bedömning av skriftliga uppgifter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Portfolio, samlar elevers skrivna texter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Annat:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Kommentar

25. Du har just svarat på en fråga om bedömningsformer, använder du samma bedömningsformer för nyanlända elever som för övriga, om du har blandade grupper? Välj ett alternativ.

- ☐ Ja, jag använder exakt samma
- ☐ Nej, jag använder annat bedömningsmaterial för de nyanlända
- ☐ Jag använder delvis samma och alla nyanlända får samma uppgifter
- ☐ Jag använder delvis samma och individanpassar för de nyanlända utifrån språklig nivå
- ☐ Jag har inga blandade grupper utan endast nyanlända

Kommentar

26. Vid betygsättning av nyanländas ämneskunskap, vilken bedömningsform anser du väger tyngst? Välj ett alternativ.

- ☐ Skriftliga prov med frågor som ska besvaras
- ☐ Muntliga prov, istället för skriftliga
- ☐ Skriftliga inlämningsuppgifter
- ☐ Muntliga redovisningar
- ☐ Annat

Kommentar

27. Vilken bedömningsform anser du generellt väger tyngst vid betygsättning? Välj ett alternativ.

- ☐ Skriftliga prov med frågor som ska besvaras
- ☐ Muntliga prov, istället för skriftliga
- ☐ Skriftliga inlämningsuppgifter
- ☐ Muntliga redovisningar
- ☐ Annat

Kommentar

28. Hur viktigt är följande vid bedömning av samtliga elevers ämneskunskap i ditt ämne/dina ämnen på en skala från mycket viktigt till inte viktigt?

Att eleven kan använda

	Mycket viktigt	Ganska viktigt	Inte viktigt
de ämnesspecifika begreppen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ämnesrelaterade förmågor, t ex argumentera, dra slutsatser	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
en ämnesspecifik genre, t ex laborationsrapport, historisk krönika	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
allmänt ämnesövergripande skolspråk	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Kommentar

29. Hur ofta använder nyanlända elever följande språk vid bedömning i ditt ämne/dina ämnen? Svara på samtliga på en skala från alltid till aldrig.

	Alltid	Ofta	Ibland	Sällan	Aldrig
Svenska	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Elevens modersmål	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Engelska	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Elevens starkaste språk, annat än de ovan	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eleven använder flera språk, beroende på vad som fungerar bäst	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eleven får rita eller använda andra uttrycksformer, t ex visa genom laboration	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Kommentar

30. Hur ofta gör du följande med dina nyanlända elever? Svara på en skala från alltid till aldrig.

	Alltid	Ofta	Ibland	Sällan	Aldrig
Ger återkoppling på/korrigerar elevens språk i klassrummet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ger återkoppling på/korrigerar elevens språk i skrivuppgifter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Rättar elevens ordval i bedömningsuppgifter/prov	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Rättar elevens språk, såsom stavning och grammatik i bedömningsuppgifter/prov	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Kommentar

Del 5. Anpassningar, stöttning och kollegialt arbete

31. Vilken typ av anpassningar gör/gör du inte när du konstruerar uppgifter i bedömningssyfte för nyanlända? Svara på samtliga på en skala från alltid till aldrig.

	Alltid	Ofta	Ibland	Sällan	Aldrig
Uppgifter översätts till elevens modersmål/starkaste språk	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ordlista/ordförklaringar på svenska tas med i uppgifterna	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ordlista/ordförklaringar på annat språk tas med i uppgiften	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Uppgifter med förenklat språk används	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Startord eller startmeningar används i frågor/uppgifter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Modelltexter eller skrivmallar används	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Flervalsfrågor eller uppgifter där eleven inte behöver skriva så mycket används i större utsträckning	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Friare uppgifter där eleven får uttrycka sig själv används i större utsträckning	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Bilder eller annat visuellt stöd används	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Annat	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Kommentar

32. Hur förbereder du eleverna inför en bedömningsuppgift? Svara på samtliga på en skala från alltid till aldrig.

	Alltid	Ofta	Ibland	Sällan	Aldrig
Går igenom typiska ämnesord och begrepp	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Går igenom användbara fraser och formuleringar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Studerar typiska texters struktur och språkliga karaktärsdrag tillsammans	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Gör skriftliga exempeluppgifter och modeller för hur man skriver olika typer av texter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Skriver texter tillsammans på lektionen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Gör exempel på muntliga mallar, t ex hur man argumenterar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Annat	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Kommentar

33. Vilken typ av anpassningar gör/gör du inte vid bedömningstillfället? Svara på samtliga på en skala från alltid till aldrig.

	Alltid	Ofta	Ibland	Sällan	Aldrig
Eleven får ta hjälp av lexikon eller Google translate	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eleven får ta hjälp av en tolk eller modersmålstalare	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eleven får ta hjälp av annan flerspråkig lärare	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eleven får svara på sitt modersmål/starkaste språk	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eleven kan använda flera olika språk, så länge han/hon gör sig förstådd	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eleven får mer tid	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Annat	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Kommentar

34. Hur gör ni med nationella prov och nyanlända elever på din skola? Ett eller flera alternativ.

Nyanlända elever

- ☐ gör inte nationella prov
- ☐ gör proven som vanligt
- ☐ gör vissa delar av nationella prov
- ☐ får välja om de vill göra proven
- ☐ får mer tid
- ☐ får ord och/eller frågor översatta av modersmålslärare eller annan flerspråkig lärare
- ☐ får använda lexikon
- ☐ får göra en engelsk version av provet

Kommentar

35. Vilka lokala riktlinjer finns för bedömning och prov med nyanlända elever på din skola?
Svara på samtliga utifrån vad som gäller på din skola.

	Uttalade riktlinjer finns mellan oss kollegor	Uttalade riktlinjer finns från skolledning/rektor	Uttalade riktlinjer kommer från kommunala tjänstemän eller förvaltning	Uttalade riktlinjer finns	Inga riktlinjer finns
Bedömning ska alltid ske på elevens modersmål/starkaste språk, exempelvis med hjälp av tolk	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Bedömning kan ske på elevens modersmål/starkaste språk, i de fall någon lärare kan det språket	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Prov gjorda på elevens modersmål/starkaste språk rättas av modersmålslärare, eller annan flerspråkig lärare	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Prov gjorda på elevens modersmål/starkaste språk översätts till svenska och rättas av mig	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Bedömning ska alltid ske på svenska, oberoende av elevens språkliga nivå	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Bedömning kan ske på engelska	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Kommentar

36. Samarbetar du eller har dialog med någon kollega för planering av undervisning och/eller bedömning av nyanlända elevers ämneskunskap? Välj ett alternativ.

- ☐ Alltid
☐ Ofta
☐ Ibland
☐ Sällan
☐ Aldrig

Kommentar

37. Med vem och på vilket sätt samarbetar du? Kryssa endast för det som är relevant för dig. Inga, ett eller flera alternativ är möjliga på varje.

	Studie- handledare	Modersmåls- lärare	SvA-lärare/Sfi- lärare/ svensklärare	Specialpedagog	Ämneskollega
Vi planerar undervisningen tillsammans	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kollegan hjälper mig med tips hur jag kan arbeta med ämnesspråket i klassrummet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kollegan hjälper mig med tips hur jag kan arbeta med det allmänna skolspråket i klassrummet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kollegan granskar språket i instruktioner, frågor och uppgifter	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kollegan hjälper eleverna med ordlistor på deras språk	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kollegan hjälper eleverna att läsa och/eller skriva texter	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kollegan hjälper till att bedöma språket i bedömningsuppgifter	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vi arbetar med samsbedömning	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Annat	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Kommentar

Del 6. Avslutande frågor

Språkutvecklande arbetssätt innebär att man på ett aktivt och medvetet sätt arbetar för att lära ut ämnesspecifikt svenska i ämneskurserna. Språkinriktad undervisning kan utformas på många sätt, men ser språket som verktyg för lärande i klassrummet.

38. Under din lärarutbildning, ingick några moment som behandlade språkutvecklande arbetssätt i förhållande till ditt ämne/dina ämnen?

- ☐ Praktiskt taget inga
☐ Några enstaka
☐ Ganska många
☐ Vet ej

Kommentar

39. Har du fått kompetensutveckling i språkutvecklande arbetssätt eller liknande under din tid som yrkesverksam lärare?

- ☐ Ja, ganska mycket
☐ Ja, vid något enstaka tillfälle
☐ Nej, ingen

Om ja, vilken sorts kompetensutveckling?

40. Är svenska ditt modersmål?

Ibland kan det finnas två modersmål/första språk, i så fall kan det anges nedan.

- ☐ Ja
☐ Nej

Om nej, eller om ytterligare modersmål finns, vilket är ditt/dina modersmål?

41. Hur skulle du bedöma dina språkkunskaper i de språk som är relevanta för dig? Kryssa endast i de som är relevanta, hoppa över de andra. Avancerad nivå motsvarar modersmåls- eller nära modersmålsnivå.

	Grundläggande nivå	Självständig nivå (medelnivå)	Avancerad nivå
Svenska	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Arabiska	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Engelska	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Finska	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Spanska	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Annat 1:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Annat 2:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Annat 3:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Om annat, ange vilket/vilka:

42. Tycker du att det är svårt med bedömning av nyanländas kunskap i ditt ämne?

- ☐ Ja
☐ Nej
☐ Ibland

Här kan du ange vad du tycker är svårt/svårast:

43. Känner du behov av kompetensutveckling kring bedömning bland nyanlända?

- ☐ Ja
☐ Nej
☐ Delvis

Om ja, vilken sorts kompetensutveckling har du störst behov av?

44. Om du vill, kan du här dela med dig av något gott råd för bedömning av nyanlända elevers ämneskunskap:

45. Skulle du vara intresserad av att bli kontaktad för att berätta mer om hur du gör vid bedömning i ditt ämne bland nyanlända elever och andraspråksinlärare?

- ☐ Ja
☐ Nej

Skriv in din mailadress/kontaktuppgifter här:

46. Här kan du lämna synpunkter eller kommentera om det finns något du tycker enkäten har missat att ta upp.

Detta avslutar enkäten. **Stort tack** för din medverkan!

Appendix D: Samtyckesblankett intervju på svenska

(original)



2019.04.23

Samtycke för personuppgiftsbehandling i forskningsstudie

Jag heter Helena Reierstam och är doktorand vid institutionen för pedagogik och didaktik vid Stockholms universitet. Jag hoppas att du vill vara med i en studie som undersöker lärares erfarenheter av bedömning av ämneskunskap bland nyanlända elever från årskurs 7 och uppåt. Syftet med studien är att förstå mer om hur lärare förhåller sig till bedömning bland nyanlända elever. Resultaten kan bli till hjälp vid utbildning av framtida ämneslärare som ska verka i flerspråkig miljö samt i diskussioner rörande aktuell policy och riktlinjer kring bedömning i den samma. För att kunna göra studien behöver jag samla in information från verksamma lärare. Datainsamlingen pågår under perioden 2018-2019. Under hösten 2018 genomfördes en enkätstudie och nu vill jag genom en uppföljande intervju med några av er som svarade på enkäten få fördjupad förståelse för era erfarenheter av bedömning bland nyanlända. Vid intervjutillfället som beräknas ta 30-40 minuter, önskar jag att du tar med dig några exempel på skriftliga bedömningsuppgifter som du har använt dig av i någon av dina kurser och som vi kan titta på.

Studien är en del av vad som kommer bli en doktorsavhandling och en av Stockholms universitet utsedd handledare stöttar mig som doktorand genom hela arbetet så att det följer alla regler. När avhandlingen är klar kommer den att diskuteras vid en offentlig doktorsdisputation och bedömas av en betygsnämnd.

Med detta samtyckesformulär kan du ge mig tillstånd att behandla information för avhandlingen. *Det är alltid frivilligt att delta.* För att kunna samla in information till studien behöver jag ditt underskrivna samtycke på den här blankettens sida 2. Om du säger ja nu kan du ändå när som helst dra tillbaka ditt samtycke och avsluta din medverkan, utan att ange orsak.

Dina personuppgifter är skyddade och kommer inte att lämnas ut till obehöriga. Inspelningar och andra uppgifter kommer att förvaras på ett säkert sätt. Den information som samlas in med hjälp av ljudinspelningar kommer snarast att göras om till anonymiserade texter så att ingen kan lista ut att just du varit med i studien. Tillståndsblanketterna kommer att förvaras i låsta utrymmen på Stockholms universitet så att de inte kan kopplas ihop med det vi spelat in.

Insamlat material kommer endast att användas i forskningssyfte och presenteras i avhandling, artiklar och vid konferenser. Studiens resultat kommer att publiceras på ett sätt så att det inte går att känna igen vilka som varit med i studien. Studien följer forskningsetiska riktlinjer och allmänna lagar. Lite längre ner på sid 2 kan du läsa mer om detta.

Det är mycket värdefullt om vi kan få ditt samtycke att genomföra studien. Hör gärna av dig om du vill fråga något.

Huvudhandledare: Ulf Fredriksson

E-post: ulf.fredriksson@edu.su.se

Telefon:

Doktorand: Helena Reierstam

E-post: helena.reierstam@edu.su.se

Telefon:

Samtycke

Jag har tagit del av information om studien och godkänner att materialet spelas in, sparas och används i forskningssyfte och presenteras i en doktorsavhandling, i artiklar och vid konferenser.

- ☐ Ja
- ☐ Nej

Informantens datum och underskrift:

Namnförtydligande:

Mera om riktlinjer och lagar för studien

Personuppgifter som är nödvändiga för att genomföra studien behandlas med stöd av samtyckeskrevet i Dataskyddsförordningen. Stockholms universitet är personuppgiftsansvarig. Enligt Dataskyddsförordningen från och med den 25 maj 2018 så har du rätt att gratis få ta del av samtliga uppgifter om dig som hanteras och vid behov få eventuella fel rättade. Du har även rätt att begära radering, begränsning eller att invända mot behandling av personuppgifter, och det finns möjlighet att inge klagomål till dataskyddsombudet på Stockholms universitet dpo@su.se eller till datainspektionen <https://www.datainspektionen.se/kontakta-oss/>

I övrigt, vid frågor vänd dig till handledare och doktorand ovan.

Appendix E: Letter of consent in English

(translated from Swedish original)

Information letter and informed consent form

My name is Helena Reierstam and I am currently a doctoral student at the Department of education at Stockholm University. My hope is that you would consider participating in my dissertation project on teachers' experiences of assessment among newly arrived students from grade 7 and up. The purpose of the study is to investigate how teachers deal with assessment among newly arrived students. The results can be instrumental in future education of subject matter teachers working in multilingual environments and contribute to discussions about current policy and guidelines for assessment. In order to conduct the study, I will need to collect data from teachers in the field. The data collection procedure takes place during fall 2018 to spring 2019. A survey was conducted during the fall 2018 and now I intend to do follow up interviews with some of you who took part in the survey to gain deeper understanding of your experiences from assessment of newly arrived students. During the interview that will last approximately 30-40 minutes, I would like for you to bring a couple of written assessment samples that have been used in some of your courses for us to look at together.

The study is part of a doctoral dissertation and a supervisor, appointed by Stockholm University, is reviewing this to make sure that the project abides by the rules. When the dissertation is finished it will be discussed at an official defense and be assessed by an examination committee.

By signing this form of consent, you give me the permission to process the data for the dissertation. Participation is always voluntary. In order to collect data for the study, I need your written consent at the back of this form, page 2. Even if you agree to participate you have the right to withdraw your consent at any time and will then be excluded from the study without further explanation.

Everyone included in the study will remain anonymous and data will not be handed over to a third party. Recordings and other data will be kept safe. Recorded material will be promptly transformed into anonymized texts that will prevent that the recordings could be traced back to you. The consent forms will be stored in a safe at Stockholm University and cannot be linked to the recordings.

Collected material will be used solely for research purposes and presented in the dissertation, articles and at research conferences. The findings will be presented in such a way that the identity of the participants cannot be revealed. The study follows the guidelines set forth by the Swedish Research Council. You can read more about this at the bottom of page 2.

I appreciate your consent to participate in the study. Please do not hesitate to contact me if you have any questions.

Main supervisor: Ulf Fredriksson

E-mail: ulf.fredriksson@edu.su.se

Telephone:

Doctoral student: Helena Reierstam

E-mail: helena.reierstam@edu.su.se

Telephone:

Consent

I have been informed about the research study and give my consent that the material is recorded, stored and used for the research purpose and presented in a dissertation, in articles and at conferences.

☐ Yes

☐ No

Date and signature of the participant:

Printed name:

More about the guidelines and laws for the study:

Personal data necessary for the study will be treated in agreement with the consent claim in the General Data Protection Regulation (GDPR). Stockholm University is controller. According to GDPR you have the right to be presented with all personal data that is processed and make corrections if called for. You also have the right to ask for deletion, limitations or object to treatment of personal data and it is possible to make complaints to the Data Protection Officer (DPO) at Stockholm University dpo@su.se or to the Swedish Data Protection Authority (DPA) <https://www.datainspektionen.se/kontakta-oss/>

For all other question, please turn to the supervisor and doctoral student above.

Appendix F: Exempel e-mail till lärare inför intervju

med förfrågan om att ta med bedömningsexempel (original)

Vid vårt samtal nämndes om du kunde ta med några prov/bedömningsexempel som du använt i någon kurs, digitalt eller på papper. I en tidigare del av min studie tog lärare med upp till fyra skriftliga prov som använts i en kurs. Vi kommer samtala om exempelvis följande:

Vad bedöms i den aktuella uppgiften/det aktuella provet?

Hur fungerade provet för de nyanlända?

Gjordes anpassningar vid provtillfället, i så fall vilka?

Hur var det att bedöma uppgiften för dig som lärare? Erfarenheter/lärdomar?

Hör av dig om du har frågor.

Med vänlig hälsning,

Helena

Appendix G: Example of Email sent to teachers before the interview

with a request to bring assessment samples (translated from the original above)

In our conversation over the phone we talked about the possibility of you bringing some exams/assessment samples that you have used in any of your courses, digitally or on paper. In my previous study, teachers brought up to four written assessment samples that had been used in a course. We will, for example, talk about the following:

What was assessed in the assignment/the exam?

How did the assessment work for the newly-arrived students?

Were accommodations used during the time of the assessment? If so, which ones?

How did it work for you as a teacher to assess the assignment? Experiences/insights?

Please let me know if you have any questions.

Kind regards,

Helena

Appendix H: Table distribution of teachers across school types and subjects

Combined distribution of teachers representing different subjects and school levels

			School level							Total
			7–9	Intro	7–9 & Intro	10–12	7–9 & 10–12	Intro & 10–12	7–9 & Intro & 10–12	
Subject(s)	NS	Count	51	12	1	14	0	3	1	83
		%	61.4%	14.5%	1.2%	16.9%	0.0%	3.6%	1.2%	100.0%
	SS	Count	36	8	1	7	2	2	0	56
		%	64.3%	14.3%	1.8%	12.5%	3.6%	3.6%	0.0%	100.0%
	Lang	Count	3	1	0	2	0	3	0	10
		%	30.0%	10.0%	0.0%	20.0%	0.0%	30.0%	0.0%	100.0%
	NS & Lang	Count	1	1	0	0	0	0	0	2
		%	50.0%	50.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%
	SS & Lang	Count	16	11	1	4	0	6	0	38
		%	42.1%	28.9%	2.6%	10.5%	0.0%	15.8%	0.0%	100.0%
	NS & SS & Lang	Count	0	1	0	0	0	1	0	2
		%	0.0%	50.0%	0.0%	0.0%	0.0%	50.0%	0.0%	100.0%
Total		Count	107	34	3	27	2	15	1	191
		%	56.0%	17.8%	1.6%	14.1%	1.0%	7.9%	0.5%	100.0%

Appendix I: Tables with results from non-parametric tests for NS and SS teachers

Presentation questions 19–33. Significant results ($< .05$) highlighted.

Question 19 <i>Scale: Never–Often</i> <i>How often did you let the students:</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...solve a short assignment individually during class?	NS	83	85.00	4 233.000	.240
	SS	94	92.53		
...solve a short assignment together in a group during class?	NS	84	86.62	4 106.000	.474
	SS	93	91.15		
...do a longer written assignment individually over a few weeks?	NS	81	84.18	4 035.500	.371
	SS	93	90.39		
...do a longer written assignment in a group over a few weeks?	NS	80	82.29	4 097.000	.180
	SS	91	91.05		
...do a prepared oral assignment individually?	NS	81	80.90	4 220.000	.098
	SS	92	92.37		
...do a prepared oral assignment in a group?	NS	80	77.81	4 295.500	.023
	SS	91	93.20		

Question 20 <i>Scale: Easy–difficult</i> <i>How difficult do you think the following is for your NAS:</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...Oral discussions in class	NS	83	93.17	3 388.500	.170
	SS	92	83.33		
... Oral presentation after preparation	NS	82	95.99	3 076.000	.025
	SS	92	79.93		
...Short written assignments	NS	83	89.43	3 699.500	.703
	SS	92	86.71		
...Longer written assignments	NS	82	92.50	3 362.000	.158
	SS	92	83.04		

Question 21 <i>Scale: Easy–Difficult</i> <i>How difficult do you think the following is for your NAS:</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...Subject-specific vocabulary and concepts	NS	83	86.75	4 004.500	.637
	SS	93	90.06		
...Subject-specific skills	NS	83	89.22	3 800.000	.847
	SS	93	87.86		
...Subject-specific oral genre	NS	81	91.95	3 325.000	.179
	SS	92	82.64		
...Subject-specific written genre	NS	80	90.72	3 422.000	.324
	SS	93	83.80		
...General academic language	NS	83	93.55	3 440.000	.183
	SS	93	83.99		
...Everyday language	NS	83	98.75	3 009.000	.008
	SS	93	79.35		

Question 22 <i>Scale: Never–Always</i> <i>How often do you think your NAS have difficulty showing their subject knowledge in the following activities:</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...Argue	NS	83	93.47	3 447.000	.168
	SS	93	84.06		
...Analyze	NS	83	84.33	4 206.000	.253
	SS	93	92.23		
...Describe	NS	83	97.06	3 066.000	.016
	SS	92	79.83		
...Discuss	NS	83	95.92	3 160.500	.035
	SS	92	80.85		
...Draw conclusions	NS	83	82.51	4 257.000	.113
	SS	93	93.85		

Question 22 continued <i>Scale: Never–Always</i> <i>How often do you think your NAS have difficulty showing their subject knowledge in the following activities:</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...Formulate questions	NS	83	90.25	3 714.000	.633
	SS	93	86.94		
...Explain	NS	83	99.05	2 984.000	.005
	SS	93	79.09		
...Reason	NS	83	91.49	3 611.000	.424
	SS	93	85.83		
...Make comparisons	NS	82	94.35	3 210.000	.069
	SS	92	81.39		
...Identify	NS	83	92.45	3 449.000	.244
	SS	92	83.99		
...Motivate	NS	83	89.48	3 695.500	.695
	SS	92	86.67		
...Name	NS	82	100.22	2 811.000	.001
	SS	93	77.23		
...Evaluate	NS	81	90.33	3 537.000	.445
	SS	93	85.03		
...Show connections	NS	83	84.39	4 118.000	.325
	SS	92	91.26		
...State	NS	83	96.99	2 989.000	.011
	SS	91	78.85		

Question 24 <i>Scale: Never–Always</i> <i>How often do you use the following with your NAS:</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...self-developed tests?	NS	81	82.36	4 263.500	.125
	SS	94	92.86		
...tests developed by colleagues?	NS	80	90.03	3 397.500	.365
	SS	92	83.43		
...standardized tests/old national tests?	NS	82	101.76	2 766.500	.001
	SS	94	76.93		
...the current national tests?	NS	81	97.81	2 850.500	.006
	SS	92	77.48		
...digital tests?	NS	82	78.99	4 552.000	.022
	SS	93	95.95		
...oral assignments?	NS	81	78.32	4591.000	.013
	SS	94	96.34		
...written assignments?	NS	80	77.03	4 517.500	.007
	SS	93	95.58		
...a portfolio?	NS	81	79.63	4 485.000	.036
	SS	94	95.21		
...other	NS	38 ⁴	28.34	424.000	.455
	SS	20	31.70		

⁴ The response rate is lower for the category *other*; however, more NS teachers (almost 50%) feel a need to report the use of other formats.

Question 28 <i>Scale: Not important–Very important</i> <i>How important is it that the students can use:</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...subject-specific concepts?	NS	82	88.71	3 590.500	.617
	SS	91	85.46		
...subject skills, e.g., to argue, draw conclusions?	NS	81	90.59	4 407.000	.022
	SS	94	94.38		
...a subject-specific genre?	NS	80	106.85	2 132.000	.000
	SS	93	69.92		
...cross-disciplinary academic language?	NS	81	80.17	4 441.000	.032
	SS	94	94.74		

Question 29 <i>Scale: Never–Always</i> <i>How often do the students use the following language in assessment?</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...Swedish	NS	81	81.94	4 054.500	.207
	SS	91	90.55		
...The student's L1/mother tongue	NS	79	77.43	3 758.000	.172
	SS	85	87.21		
...English	NS	79	81.07	3 470.500	.689
	SS	85	83.83		
...The student's strongest language	NS	74	76.03	3 290.500	.406
	SS	83	81.64		
...The student uses a mix of languages, depending on what works best	NS	78	85.63	3 265.500	.578
	SS	88	81.61		
...Draw or use other forms of expression	NS	81	99.44	2 151.500	.000
	SS	85	68.31		

Question 30 <i>Scale: Never–Always</i> <i>How often do you do the following with your NAS:</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
... Give corrective feedback on language in class?	NS	80	83.41	4 007.500	.363
	SS	93	90.09		
... Give corrective feedback on language in written assignments?	NS	80	83.18	3 945.500	.398
	SS	92	89.39		
... Correct the student's choice of words in tests/assignments used for assessment?	NS	80	81.62	4 070.500	.210
	SS	92	90.74		
... Correct the student's language (e.g., spelling and grammar) in tests/assignments used for assessment?	NS	80	79.51	4 319.000	.059
	SS	93	93.44		

Question 31 <i>Scale: Never–Always</i> <i>What type of accommodations do you use when you develop tests/assignments used for assessment for NAS?</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
... Assignments are translated into the student's L1/strongest language	NS	80	85.92	3 726.500	.881
	SS	92	87.01		
... Wordlists/explanations of words in Swedish	NS	80	77.19	4 425.000	.018
	SS	92	94.60		
... Wordlists/explanations of words in another language	NS	80	80.97	3 882.500	.282
	SS	89	88.62		
... Assignments with simplified language	NS	80	82.33	4 093.500	.239
	SS	93	91.02		
... Sentence starters	NS	76	69.26	4 730.000	.000
	SS	93	97.86		
... Model texts	NS	78	70.22	4 780.000	.000
	SS	92	98.46		

Question 31 continued <i>Scale: Never–Always</i> <i>What type of accommodations do you use when you develop tests/assignments used for assessment for NAS?</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...Multiple-choice questions or questions requiring less writing	NS	79	81.66	3 977.000	.265
	SS	92	89.73		
...Open questions where the students can express themselves	NS	78	72.91	4 648.000	.001
	SS	93	96.98		
...Pictures or other visual support	NS	78	83.96	3 630.500	.784
	SS	91	85.90		
...Other	NS	34	27.82	431.000	.501
	SS	23	30.74		

Question 32 <i>Scale: Never–Always</i> <i>How do you prepare the students for an assessment?</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...Look at subject-specific vocabulary and concepts	NS	80	80.91	4 287.000	.031
	SS	94	93.11		
...Look at useful sentence structures	NS	80	86.94	3 645.000	.910
	SS	92	86.12		
...Study the structure of typical texts together	NS	80	80.89	4 208.500	.126
	SS	93	92.25		
...Provide written examples/model texts	NS	80	84.66	3 987.500	.475
	SS	94	89.92		
...Write texts together in class	NS	80	74.90	4 608.000	.003
	SS	92	96.59		
...Provide oral models	NS	80	73.41	4 807.500	.001
	SS	93	98.69		
...other	NS	27	21.09	321.500	.131
	SS	19	26.92		

Question 33 <i>Scale: Never–Always</i> <i>What type of accommodations do you use in the assessment situation?</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...The student can use a dictionary or Google Translate.	NS	81	86.78	3 581.500	.838
	SS	90	85.29		
...The student can use an interpreter.	NS	80	91.14	3 229.000	.192
	SS	91	81.48		
...The student can receive help from another bilingual teacher.	NS	79	86.65	3 345.500	.577
	SS	89	82.59		
...The student can answer in his/her L1/strongest language.	NS	81	87.64	3 674.000	.870
	SS	92	86.43		
...The student can use a mix of languages as long as he/she makes him/herself understood.	NS	81	88.69	3 670.000	.766
	SS	93	86.46		
...The student gets extended time.	NS	80	85.78	3 817.500	.752
	SS	93	88.05		
...Other.	NS	25	17.76	156.000	.542
	SS	11	20.18		

Appendix J: Tables with results from non-parametric tests for SS and SSLang teachers

Presentation questions 19–33. Significant results (< .05) highlighted.

Question 19 <i>Scale: Never–Often</i> <i>How often do you let the students:</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...solve a short assignment individually during class?	SS	59	52.54	1 030.000	.183
	SSLang	40	46.25		
...solve a short assignment together in a group during class?	SS	58	49.03	1 187.000	.809
	SSLang	40	50.17		
...do a longer written assignment individually over a few weeks?	SS	58	46.54	1 331.500	.173
	SSLang	40	53.79		
...do a longer written assignment in a group over a few weeks?	SS	58	46.38	1 341.000	.135
	SSLang	40	54.02		
... do a prepared oral assignment individually?	SS	58	43.14	1 471.000	.006
	SSLang	39	57.72		
...do a prepared oral assignment in a group?	SS	56	45.93	1 264.000	.238
	SSLang	40	52.10		

Question 20 <i>Scale: Easy–Difficult</i> <i>How difficult do you think the following is for your NAS:</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...Oral discussions in class	SS	58	50.47	1 046.000	.506
	SSLang	39	46.82		
... Oral presentation after preparation	SS	58	53.82	851.500	.031
	SSLang	39	41.83		
...Short written assignments	SS	58	48.27	1 173.500	.736
	SSLang	39	50.09		
...Longer written assignments	SS	58	48.59	1 154.500	.848
	SSLang	39	49.60		

Question 21 <i>Scale: Easy–Difficult</i> <i>How difficult do you think the following is for your NAS:</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...Subject-specific vocabulary and concepts	SS	58	50.03	1 129.000	.806
	SSLang	40	48.73		
...Subject-specific skills	SS	58	53.24	943.000	.087
	SSLang	40	44.08		
...Subject-specific oral genre	SS	57	53.14	904.000	.052
	SSLang	40	43.10		
...Subject-specific written genre	SS	58	53.34	937.000	.079
	SSLang	40	43.92		
...General academic language	SS	58	48.22	1 234.000	.565
	SSLang	40	51.35		
...Everyday language	SS	58	54.38	877.000	.031
	SSLang	40	42.42		

Question 22 <i>Scale: Never–Always</i> <i>How often do you think your NAS have difficulty showing their subject knowledge in the following activities:</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...Argue	SS	58	52.83	967.000	.119
	SSLang	40	44.67		
...Analyze	SS	58	46.59	1 329.000	.180
	SSLang	40	53.73		
...Describe	SS	58	48.15	1 180.500	.695
	SSLang	39	50.27		
...Discuss	SS	58	49.89	1 079.500	.682
	SSLang	39	47.68		
...Draw conclusions	SS	58	48.50	1 218.000	.653
	SSLang	40	50.95		

Question 22 continued <i>Scale: Never–Always</i> <i>How often do you think your NAS have difficulty showing their subject knowledge in the following activities:</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...Formulate questions	SS	58	49.98	1 132.000	.823
	SSLang	40	48.80		
...Explain	SS	58	51.33	1 054.000	.398
	SSLang	40	46.85		
...Reason	SS	58	48.40	1 224.000	.618
	SSLang	40	51.10		
...Make comparisons	SS	58	51.55	983.000	.245
	SSLang	39	45.21		
...Identify	SS	57	50.76	1 039.500	.439
	SSLang	40	46.49		
...Motivate	SS	57	49.34	1 120.500	.880
	SSLang	40	48.51		
...Name	SS	58	48.42	1 222.500	.626
	SSLang	40	51.06		
...Evaluate	SS	58	48.16	1 237.500	.546
	SSLang	40	51.44		
...Show connections	SS	57	49.47	1 113.000	.828
	SSLang	40	48.33		
...State	SS	57	49.08	1 078.5000	.792
	SSLang	39	47.65		

Question 24 <i>Scale: Never-Always</i> <i>How often do you use the following for assessment:</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...self-developed tests?	SS	59	51.71	1 079.000	.429
	SSLang	40	47.48		
...tests developed by colleagues?	SS	59	48.19	1 169.000	.712
	SSLang	38	50.26		
...standardized tests/old national tests?	SS	59	43.07	1 589.000	.002
	SSLang	40	60.23		
...the current national tests?	SS	57	46.06	1 307.500	.205
	SSLang	40	53.19		
...digital tests?	SS	58	49.52	1 159.000	.994
	SSLang	40	49.48		
....oral assignments?	SS	59	48.47	1 270.500	.492
	SSLang	40	52.26		
...written assignments?	SS	59	48.88	1 187.000	.767
	SSLang	39	50.44		
...a portfolio?	SS	59	45.09	1 469.500	.034
	SSLang	40	57.24		
...other?	SS	13	11.65	56.500	.890
	SSLang	9	11.28		

Question 28 <i>Scale: Very important–Not important</i> <i>How important is it that the students are able to use:</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...subject-specific concepts?	SS	58	50.17	1 005.000	.396
	SSLang	38	45.95		
...subject-specific skills?	SS	59	50.93	1 125.000	.586
	SSLang	40	48.62		
...subject-specific genre?	SS	58	51.37	1 051.500	.374
	SSLang	40	46.79		
...cross-disciplinary academic language?	SS	59	45.73	1 432.000	.043
	SSLang	40	56.30		

Question 29 <i>Scale: Never–Always</i> <i>How often do the students use the following language in assessment?</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...Swedish	SS	57	47.71	1 156.500	.703
	SSLang	39	49.65		
...The student's L1/mother tongue	SS	51	42.39	1 051.000	.347
	SSLang	37	47.41		
...English	SS	54	42.08	1 102.500	.169
	SSLang	35	49.50		
...The student's strongest language	SS	52	42.13	955.000	.503
	SSLang	34	45.59		
...The student uses a mix of languages depending on what works best	SS	54	43.73	1 121.500	.306
	SSLang	37	49.31		
...Draw or use other forms of expression	SS	53	45.62	921.000	.776
	SSLang	36	44.08		

Question 30 <i>Scale: Never–Always</i> <i>How often do you do the following with your NAS:</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
... Give corrective feedback on language in class?	SS	58	48.84	1 198.000	.775
	SSLang	40	50.45		
... Give corrective feedback on language in written assignments?	SS	57	45.45	1 342.500	.124
	SSLang	40	54.06		
... Correct the student's choice of words in tests/assignments used for assessment?	SS	57	48.70	1 157.000	.896
	SSLang	40	49.42		
... Correct the student's language (e.g., spelling and grammar) in tests/assignments used for assessment?	SS	58	49.41	1 165.000	.970
	SSLang	40	49.62		

Question 31 <i>Scale: Never–Always</i> <i>What type of accommodations do you use when you develop tests/assignments used for assessment for NAS?</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...Assignments are translated into the student's L1/strongest language	SS	57	50.97	970.500	.271
	SSLang	39	44.88		
... Wordlists/explanations of words in Swedish	SS	57	43.75	1 382.500	.036
	SSLang	39	55.45		
... Wordlists/explanations of words in another language	SS	55	47.78	1 002.000	.725
	SSLang	38	45.87		
...Assignments with simplified language	SS	58	53.00	899.000	.077
	SSLang	39	43.05		
...Sentence starters	SS	58	46.60	1 270.000	.293
	SSLang	39	52.56		
...Model texts	SS	58	42.11	1 472.500	.004
	SSLang	38	58.25		

Question 31 Continued <i>Scale: Never–Always</i> <i>What type of accommodations do you use when you develop tests/assignments used for assessment for NAS?</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...Multiple-choice questions or questions requiring less writing	SS	57	49.22	1 070.500	.747
	SSLang	39	47.45		
...Open questions where the students can express themselves	SS	58	44.87	1 370.500	.054
	SSLang	39	55.14		
...Pictures or other visual support	SS	57	47.19	1 129.000	.706
	SSLang	38	49.21		
...Other	SS	14	11.61	96.500	.291
	SSLang	11	14.77		

Question 32 <i>Scale: Never–Always</i> <i>How do you prepare the students for an assessment?</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...Look at subject-specific vocabulary and concepts	SS	59	49.81	1 191.000	.902
	SSLang	40	50.27		
...Look at useful sentence structures	SS	58	42.64	1 500.000	.004
	SSLang	39	58.46		
...Study the structure of typical texts together	SS	58	39.24	1 755.000	.000
	SSLang	40	64.38		
...Provide written examples/model texts	SS	59	40.27	1 754.000	.000
	SSLang	40	64.35		
...Write texts together in class	SS	59	39.90	1 658.000	.000
	SSLang	38	63.13		
...Provide oral models	SS	59	42.15	1 584.000	.001
	SSLang	39	60.62		
...Other	SS	14	9.50	56.000	.274
	SSLang	6	12.83		

Question 33 <i>Scale: Never–Always</i> <i>What type of accommodations do you use in the assessment situation?</i>	Categories	N	Mean rank	Mann-Whitney U	Asymp. Sig. (2-sided)
...The student can use a dictionary or Google Translate.	SS	57	50.22	956.500	.318
	SSLang	38	44.67		
...The student can use an interpreter.	SS	58	50.79	969.000	.307
	SSLang	38	45.00		
...The student can receive help from another bilingual teacher.	SS	57	48.66	988.500	.59958
	SSLang	37	45.72		
...The student can answer in his/her LI/strongest language.	SS	58	52.14	949.000	.166
	SSLang	39	44.33		
...The student can use a mix of languages as long as he/she makes him/herself understood.	SS	59	51.50	1 032.500	.382
	SSLang	39	46.46		
...The student gets extended time.	SS	59	51.15	1 053.000	.448
	SSLang	39	47.00		
...Other.	SS	8	5.38	25.000	.154
	SSLang	4	8.75		

Appendix K: Tables with results from non-parametric tests across school types

Presentation questions 19–33. Three categories: *Grades 7–9, introductory program* and *Grades 10–12*. Significant results (< .05) highlighted. *< .005 adjusted using a Bonferroni post hoc correction.

Question 19 <i>Scale: Never–Often</i> <i>How often do you let the students:</i>	Categories	N	Mean rank	Kruskal-Wallis	Asymp. Sig. (2-sided)
...solve a short assignment individually during class?	7–9	107	79.14	6.442	.040*
	Intro	34	92.10		
	10–12	28	98.77		
...solve a short assignment together in a group during class?	7–9	106	84.26	.220	.896
	Intro	35	87.87		
	10–12	28	84.21		
...do a longer written assignment individually over a few weeks?	7–9	104	80.17	1.977	.372
	Intro	35	92.00		
	10–12	27	85.31		
...do a longer written assignment in a group over a few weeks?	7–9	102	80.31	2.443	.295
	Intro	35	82.19		
	10–12	28	93.82		
...do a prepared oral assignment individually?	7–9	102	73.13	13.761	.001*
	Intro	35	99.71		
	10–12	28	98.07		
...do a prepared oral assignment in a group?	7–9	100	76.58	4.605	.100
	Intro	35	87.54		
	10–12	28	94.41		

Question 20 <i>Scale: Easy–Difficult</i> <i>How difficult do you think the following is for your NAS:</i>	Categories	N	Mean rank	Kruskal-Wallis	Asymp. Sig. (2-sided)
...Oral discussions in class	7–9	104	94.95	18.650	.000*
	Intro	34	66.85		
	10–12	28	61.18		
...Oral presentations after preparation	7–9	104	92.48	13.228	.001*
	Intro	34	70.38		
	10–12	27	62.39		
...Short written assignments	7–9	104	95.28	22.648	.000*
	Intro	34	72.97		
	10–12	28	52.54		
...Longer written assignments	7–9	104	87.93	4.976	.083
	Intro	34	79.72		
	10–12	27	68.13		

Question 21 <i>Scale: Easy–Difficult</i> <i>How difficult do you think the following is for your NAS:</i>	Categories	N	Mean rank	Kruskal-Wallis	Asymp. Sig. (2-sided)
...Subject-specific vocabulary and concepts	7–9	105	91.58	8.585	.014*
	Intro	34	69.40		
	10–12	28	73.30		
...Subject-specific skills	7–9	105	90.66	7.765	.021*
	Intro	34	78.51		
	10–12	28	65.68		
...Subject-specific oral genre	7–9	103	90.21	10.128	.006*
	Intro	33	75.03		
	10–12	28	62.93		

Question 21 <i>Continued</i> <i>Scale: Easy–Difficult</i> <i>How difficult do you think the following is for your NAS:</i>	Categories	N	Mean rank	Kruskal-Wallis	Asymp. Sig. (2-sided)
...Subject-specific written genre	7–9	103	90.65	12.450	.002*
	Intro	33	77.39		
	10–12	28	58.55		
...General academic language	7–9	105	89.98	5.041	.080
	Intro	34	75.24		
	10–12	28	72.23		
...Everyday language	7–9	105	94.04	15.222	.000*
	Intro	34	59.72		
	10–12	28	75.84		

Question 22 <i>Scale: Never–Always</i> <i>How often do you think your NAS have difficulty showing their subject knowledge in the following activities:</i>	Categories	N	Mean rank	Kruskal-Wallis	Asymp. Sig. (2-sided)
...Argue	7–9	105	93.73	19.496	.000*
	Intro	33	74.73		
	10–12	28	55.46		
...Analyze	7–9	105	87.60	7.817	.020*
	Intro	33	88.12		
	10–12	28	62.66		
...Describe	7–9	105	89.80	5.945	.051
	Intro	33	75.14		
	10–12	28	69.73		
...Discuss	7–9	105	89.90	7.056	.029
	Intro	33	70.45		
	10–12	27	71.50		

Question 22 <i>Continued</i> <i>Scale: Never–Always</i> <i>How often do you think your NAS have difficulty showing their subject knowledge in the following activities:</i>	Categories	N	Mean rank	Kruskal-Wallis	Asymp. Sig. (2-sided)
...Draw conclusions	7–9	105	87.16	4.380	.112
	Intro	33	85.36		
	10–12	28	67.57		
...Formulate questions	7–9	105	84.95	.320	.852
	Intro	33	81.38		
	10–12	28	80.55		
...Explain	7–9	105	90.46	7.152	.028
	Intro	33	70.05		
	10–12	28	73.25		
....Reason	7–9	105	92.05	10.681	.005*
	Intro	33	70.47		
	10–12	28	66.79		
...Make comparisons	7–9	104	90.22	8.854	.012*
	Intro	33	67.33		
	10–12	27	71.30		
...Identify	7–9	105	87.51	3.013	.222
	Intro	33	73.42		
	10–12	27	77.15		
...Motivate	7–9	104	89.75	6.876	.032
	Intro	33	74.45		
	10–12	28	67.98		
...Name	7–9	105	88.26	4.064	.131
	Intro	33	74.42		
	10–12	27	73.04		

Question 22 <i>Continued</i> <i>Scale: Never–Always</i> <i>How often do you think your NAS have difficulty showing their subject knowledge in the following activities:</i>	Categories	N	Mean rank	Kruskal-Wallis	Asymp. Sig. (2-sided)
...Evaluate	7–9	104	82.78	.150	.928
	Intro	32	83.92		
	10–12	28	79.84		
...Show connections	7–9	105	89.41	7.717	.021*
	Intro	32	77.46		
	10–12	28	65.16		
...State	7–9	104	89.99	6.902	.032
	Intro	33	72.23		
	10–12	28	69.73		

Question 24 <i>Scale: Never–Always</i> <i>How often do you use the following for assessment:</i>	Categories	N	Mean rank	Kruskal-Wallis	Asymp. Sig. (2-sided)
...self-developed tests?	7–9	105	80.52	2.799	.247
	Intro	35	87.84		
	10–12	28	95.23		
...tests developed by colleagues?	7–9	103	81.14	.411	.814
	Intro	34	86.88		
	10–12	27	82.17		
...standardized tests/old national tests?	7–9	106	85.41	1.057	.590
	Intro	35	89.54		
	10–12	28	77.79		
...the current national tests?	7–9	105	89.20	4.526	.104
	Intro	35	69.81		
	10–12	28	85.21		

Question 24 <i>Continued</i> <i>Scale: Never–Always</i> <i>How often do you use the following for assessment:</i>	Categories	N	Mean rank	Kruskal-Wallis	Asymp. Sig. (2-sided)
...digital tests?	7–9	106	88.53	2.437	.296
	Intro	35	74.67		
	10–12	27	81.43		
...oral assignments?	7–9	106	81.31	1.432	.489
	Intro	34	90.56		
	10–12	28	89.23		
...written assignments?	7–9	103	84.42	.139	.933
	Intro	35	81.40		
	10–12	28	82.73		
...a portfolio?	7–9	106	75.45	12.676	.002*
	Intro	34	107.59		
	10–12	28	90.73		
...other?	7–9	38	29.36	.056	.972
	Intro	11	28.27		
	10–12	8	28.31		

Question 28 <i>Scale: Very important–Not important</i> <i>How important is it that the students can use:</i>	Categories	N	Mean rank	Kruskal-Wallis	Asymp. Sig. (2-sided)
...subject-specific concepts?	7–9	104	84.38	1.479	.477
	Intro	33	75.48		
	10–12	28	86.75		
...subject-specific skills?	7–9	105	83.94	4.387	.112
	Intro	34	75.07		
	10–12	28	95.07		

Question 28 <i>Continued</i> <i>Scale: Very important–</i> <i>Not important</i> <i>How important is it that</i> <i>the students can use:</i>	Categories	N	Mean rank	Kruskal- Wallis	Asymp. Sig. (2- sided)
...subject-specific genre?	7–9	103	81.69	2.394	.302
	Intro	34	77.91		
	10–12	28	94.00		
...cross-disciplinary academic language?	7–9	105	77.52	7.996	.018*
	Intro	34	101.16		
	10–12	28	87.46		

Question 29 <i>Scale: Never–Always</i> <i>How often do the</i> <i>students use the</i> <i>following language in</i> <i>assessment?</i>	Categories	N	Mean rank	Kruskal- Wallis	Asymp. Sig. (2- sided)
...Swedish	7–9	104	68.45	31.756	.000*
	Intro	33	108.95		
	10–12	28	196.45		
...The student's L1/mother tongue	7–9	101	86.11	14.960	.001*
	Intro	28	71.00		
	10–12	25	49.98		
...English	7–9	103	79.24	.623	.732
	Intro	28	72.20		
	10–12	24	79.44		
...The student's strongest language, other than the above	7–9	98	79.85	5.421	.066
	Intro	27	71.54		
	10–12	24	59.08		

Question 29 <i>Continued</i> <i>Scale: Never–Always</i> <i>How often do the students use the following language in assessment?</i>	Categories	N	Mean rank	Kruskal-Wallis	Asymp. Sig. (2-sided)
...The student uses a mix of languages depending on what works best	7–9	102	84.49	4.717	.095
	Intro	28	71.50		
	10–12	27	66.06		
...Draw or use other forms of expression	7–9	103	81.52	.656	.720
	Intro	30	74.52		
	10–12	25	77.16		

Question 30 <i>Scale: Never–Always</i> <i>How often do you do the following with your NAS:</i>	Categories	N	Mean rank	Kruskal-Wallis	Asymp. Sig. (2-sided)
... Give corrective feedback on language in class?	7–9	103	71.93	19.244	.000*
	Intro	35	110.34		
	10–12	28	92.52		
... Give corrective feedback on language in written assignments?	7–9	103	73.88	15.281	.000*
	Intro	34	86.75		
	10–12	28	111.98		
... Correct the student's choice of words in tests/assignments used for assessment?	7–9	102	78.10	6.668	.000*
	Intro	35	81.11		
	10–12	28	103.20		
... Correct the student's language (e.g., spelling and grammar) in tests/assignments used for assessment?	7–9	103	79.32	8.081	.018*
	Intro	35	77.59		
	10–12	28	106.27		

Question 31 <i>Scale: Never–Always</i> <i>What type of accommodations do you use when you develop tests/assignments used for assessment among NAS?</i>	Categories	N	Mean rank	Kruskal-Wallis	Asymp. Sig. (2-sided)
...Assignments are translated into the student's L1/strongest language	7–9	101	96.18	26.721	.000*
	Intro	34	63.50		
	10–12	28	53.32		
...Wordlists/explanations of words in Swedish	7–9	101	78.11	2.165	.339
	Intro	34	91.04		
	10–12	28	85.04		
...Wordlists/explanations of words in another language	7–9	99	82.01	1.299	.522
	Intro	34	83.00		
	10–12	27	71.81		
...Assignments with simplified language	7–9	102	77.88	5.530	.063
	Intro	34	98.91		
	10–12	28	79.41		
...Sentence starters	7–9	100	78.22	2.156	.340
	Intro	34	90.47		
	10–12	26	76.23		
...Model texts	7–9	100	74.22	5.998	.050
	Intro	34	93.37		
	10–12	27	90.54		
...Multiple-choice questions or questions requiring less writing	7–9	102	87.21	4.489	.106
	Intro	33	72.61		
	10–12	27	70.81		
...Open questions where the students can express themselves	7–9	100	78.44	1.045	.593
	Intro	34	83.32		
	10–12	27	87.54		
...Pictures or other visual support	7–9	100	78.83	.390	.823
	Intro	33	83.20		
	10–12	27	83.37		

Question 31 Continued <i>Scale: Never–Always</i> <i>What type of accommodations do you use when you develop tests/assignments used for assessment among NAS?</i>	Categories	N	Mean rank	Kruskal-Wallis	Asymp. Sig. (2-sided)
...Other	7–9	33	26.20	.638	.727
	Intro	11	29.86		
	10–12	10	29.20		

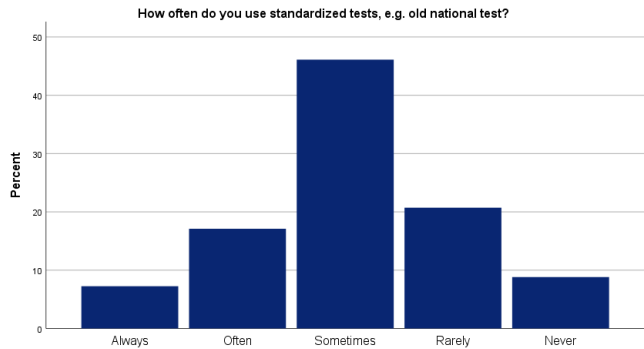
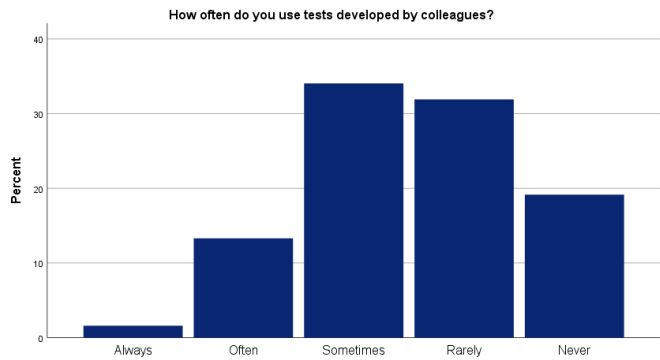
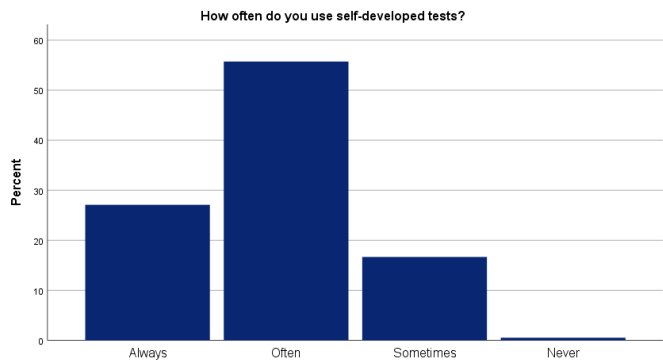
Question 32 <i>Scale: Never–Always</i> <i>How do you prepare the students for an assessment?</i>	Categories	N	Mean rank	Kruskal-Wallis	Asymp. Sig. (2-sided)
...Look at subject-specific vocabulary and concepts.	7–9	103	81.62	1.239	.538
	Intro	35	89.40		
	10–12	28	83.05		
...Look at useful sentence structures.	7–9	103	80.38	.857	.652
	Intro	34	88.63		
	10–12	27	82.87		
...Study the structure of typical texts together.	7–9	102	75.15	7.996	.018*
	Intro	35	98.77		
	10–12	28	91.88		
...Provide written examples/model texts.	7–9	103	76.39	6.407	.041
	Intro	35	95.39		
	10–12	28	94.80		
...Write texts together in class.	7–9	102	77.76	3.054	.217
	Intro	34	92.06		
	10–12	28	88.16		
...Provide oral models.	7–9	102	78.45	5.578	.061
	Intro	35	99.14		
	10–12	28	79.39		

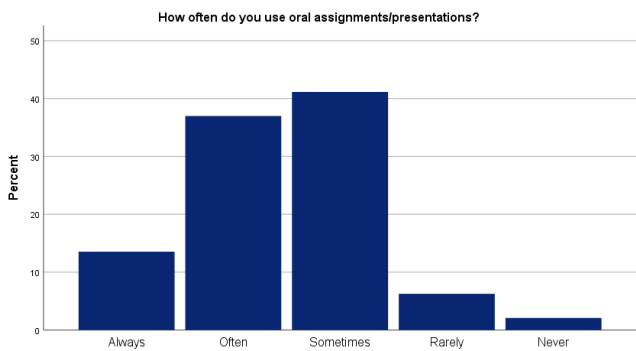
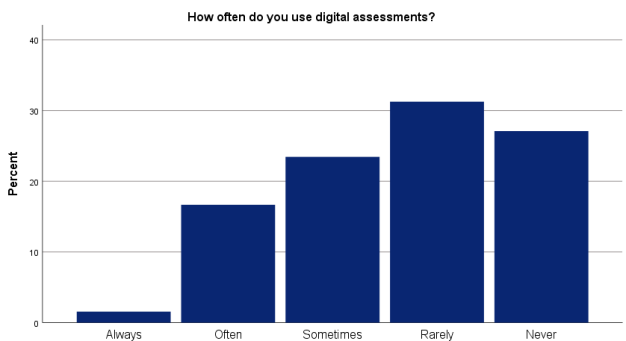
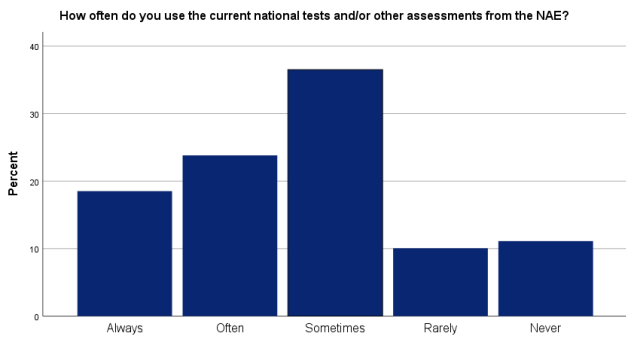
Question 32 <i>Continued</i> <i>Scale: Never–Always</i> <i>How do you prepare the students for an assessment?</i>	Categories	N	Mean rank	Kruskal-Wallis	Asymp. Sig. (2-sided)
...Other	7–9	29	21.43	1.144	.565
	Intro	6	27.33		
	10–12	9	22.72		

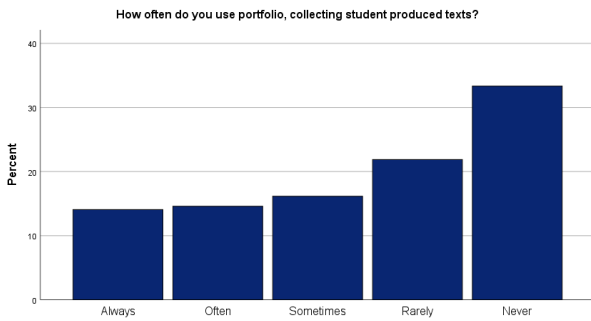
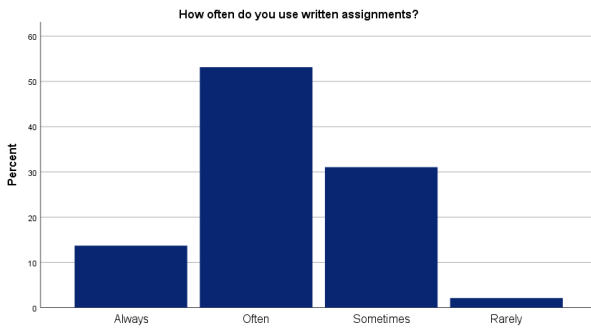
Question 33 <i>Scale: Never–Always</i> <i>What type of accommodations do you use in the assessment situation?</i>	Categories	N	Mean rank	Kruskal-Wallis	Asymp. Sig. (2-sided)
...The student can use a dictionary or Google Translate.	7–9	100	90.58	12.692	.002*
	Intro	35	58.94		
	10–12	28	80.16		
...The student can use an interpreter.	7–9	102	100.96	40.354	.000*
	Intro	35	58.16		
	10–12	28	48.64		
...The student can receive help from another bilingual teacher.	7–9	102	87.70	8.034	.018*
	Intro	35	78.37		
	10–12	28	60.52		
...The student can answer in his/her L1/strongest language.	7–9	102	97.44	25.964	.000*
	Intro	35	61.14		
	10–12	28	57.73		
...The student can use a mix of languages as long as he/she makes him/herself understood.	7–9	103	93.80	13.197	.001*
	Intro	35	64.37		
	10–12	28	69.54		

Question 33 Continued <i>Scale: Never–Always</i> <i>What type of accommodations do you use in the assessment situation?</i>	Categories	N	Mean rank	Kruskal-Wallis	Asymp. Sig. (2-sided)
...The student gets extended time.	7–9	103	86.59	1.994	.369
	Intro	34	74.62		
	10–12	28	79.96		
...Other.	7–9	20	16.68	1.009	.604
	Intro	7	20.43		
	10–12	8	19.19		

Appendix L: Bar charts: frequency use of different assessment formats







Today's educational environment is increasing multilingual where teachers and students alike are communicating and learning in multiple languages. The experience is often challenging for these students who, at the same time, are learning the language of instruction and are expected to develop subject area knowledge. Teacher beliefs about and assessment of these students' learning is thus of particular importance when considering this phenomenon and, as of yet, is a largely unexplored research area. The aim of this thesis is to contribute to this research area and draw attention to the consequences varying language policies and pedagogies may have on fairness and validity in assessment outcomes.

This thesis includes two studies on teachers' assessment beliefs and reported assessment practices in Swedish schools:

- In Content and Language Integrated Learning (CLIL) context, Grades 10-12.
- In contexts with newly-arrived migrant students (NAS), Grades 7-12.



Helena Reierstam

Taught English and French as a foreign language in Swedish upper secondary schools and ESP in Higher Vocational Education. More recently she held a position as a Dir. of Ed. at the Swedish National Agency for Education where she worked with national tests.

ISBN 978-91-7911-250-9