Occupation-Based and Occupation-Focused Evaluation and Intervention with Children

A Validation Study of the Assessment of Motor and Process Skills (AMPS)

Brigitte Elisabeth Gantschnig
To Dietmar and Johannes
# CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>VIII</td>
</tr>
<tr>
<td>ABBREVIATIONS</td>
<td>X</td>
</tr>
<tr>
<td>ORIGINAL PAPERS</td>
<td>XI</td>
</tr>
<tr>
<td>PREFACE</td>
<td>XII</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Occupational Therapy With Children and in the Context of Middle Europe</td>
<td>2</td>
</tr>
<tr>
<td>Children With and at Risk for Mild Disabilities</td>
<td>2</td>
</tr>
<tr>
<td>Occupational Therapy in the Context of Middle Europe</td>
<td>7</td>
</tr>
<tr>
<td>Evaluating ADL Task Performance</td>
<td>9</td>
</tr>
<tr>
<td>Concept of Occupational Performance</td>
<td>9</td>
</tr>
<tr>
<td>Evaluating ADL Task Performance Within the Occupational Therapy</td>
<td>10</td>
</tr>
<tr>
<td>Process</td>
<td></td>
</tr>
<tr>
<td>Evaluating ADL Task Performance Using the AMPS</td>
<td>11</td>
</tr>
<tr>
<td>Validity of Evaluation Tools</td>
<td>15</td>
</tr>
<tr>
<td>Concept of Validity</td>
<td>15</td>
</tr>
<tr>
<td>Approaches for Evaluating Validity Evidence</td>
<td>16</td>
</tr>
<tr>
<td>Evidence Supporting Validity of the AMPS</td>
<td>19</td>
</tr>
<tr>
<td>RATIONALE</td>
<td>23</td>
</tr>
<tr>
<td>AIMS OF THE THESIS</td>
<td>25</td>
</tr>
<tr>
<td>Specific Aims</td>
<td>25</td>
</tr>
<tr>
<td>METHODS</td>
<td>27</td>
</tr>
<tr>
<td>Participants</td>
<td>27</td>
</tr>
<tr>
<td>Study I</td>
<td>27</td>
</tr>
<tr>
<td>Study II</td>
<td>29</td>
</tr>
<tr>
<td>Study III</td>
<td>29</td>
</tr>
<tr>
<td>Study IV</td>
<td>30</td>
</tr>
<tr>
<td>Enrollment</td>
<td>31</td>
</tr>
<tr>
<td>Evaluation Tools and Procedures</td>
<td>32</td>
</tr>
<tr>
<td>Use of the AMPS in Studies I to IV</td>
<td>32</td>
</tr>
<tr>
<td>Use of the COPM in Study IV</td>
<td>33</td>
</tr>
<tr>
<td>Procedures in Study IV</td>
<td>34</td>
</tr>
</tbody>
</table>
Data Analysis ................................................................................. 37
  Descriptive Statistical Analyses .................................................. 37
  Rasch Analysis Methods............................................................. 37
  Inferential Statistical Analyses................................................. 38
  Effect Sizes .............................................................................. 40
  Descriptive Analysis of Feasibility ............................................. 41

Ethical Considerations ........................................................... 42

RESULTS .................................................................................. 45

Evidence Supporting Validity of the AMPS ......................... 45
  Evidence Related to Internal Structure .................................. 45
  Evidence Based on Relation to Other Variables ..................... 46
  Evidence Related to Consequences of Testing and Test Fairness 47

DISCUSSION ........................................................................ 51

Discussion of Evidence Supporting Validity of the AMPS .......... 51
  Evidence Related to Internal Structure .................................. 51
  Evidence Based on Relation to Other Variables ..................... 55
  Evidence Related to Consequences of Testing and Test Fairness 58

Implications and Recommendations for Occupational Therapy
Practice and Research .......................................................... 61
  Implications for Occupational Therapy Practice ................. 61
  Implications and Recommendations for Further Research .... 62

Other Methodological Considerations ...................................... 64
  Other Methodological Considerations Related to Effect Sizes 64
  Other Methodological Considerations Related to Study IV ..... 65

CONCLUSIONS ...................................................................... 67

ACKNOWLEDGEMENTS ........................................................ 69

REFERENCES .......................................................................... 71

APPENDICES .......................................................................... 87
Abstract

ABSTRACT

Introduction

Occupational therapists are concerned with enabling people to perform the daily life tasks they need, want, or are expected to perform for fullest possible integration into community living and participation in society. Children with mild disabilities have problems performing personal and instrumental activities of daily living (ADL) tasks at home or school, and that can limit their full integration and participation in their homes and school lives. There is a need, therefore, to identify their specific problems with ADL task performance so as to be able to develop effective interventions. Not only, there is a need for evidence related to effectiveness of occupational therapy interventions for children with mild disabilities, but also a need for valid occupational-therapy-specific evaluation tools for use with children.

Purpose

The purpose of this thesis was to contribute evidence to support the valid use of the Assessment of Motor and Process Skills (AMPS) with children, including children living in Middle Europe. More specifically, I aimed to evaluate validity evidence from different sources related to the use of the AMPS in occupation-based and occupation-focused evaluation and intervention.

Method

This thesis consisted of four studies, implemented in two phases. Phase one focused on evaluation of a) validity evidence of the AMPS scales in relation to internal structure and stability of item difficulty calibration values for a Middle European sample compared to samples from other world regions (Study I); b) the stability of the mean AMPS measures between typically-developing children from Middle Europe and from other world regions (Study II); and c) the sensitivity of the AMPS measures to discriminate between typically-developing children and children with and at risk for mild disabilities (Study III). Participants for phase one were from both Middle Europe and from other world regions and they were selected from the AMPS database, Ft. Collins, Colorado, USA. Data were analyzed using many-facet Rasch analyses, ANOVAs, regression analyses, related post-hoc tests, and effect size calculations.

Phase two of the research project focused on evaluating validity evidence for the use of the AMPS as a standardized, occupation-based, and occupation-focused evaluation tool in the context of a feasibility study with children with mild disabilities implemented in a Swiss setting (Study IV). Data were analyzed based on feasibility objectives and the principles of deductive content analysis. The evaluation of validity evidence of the
Abstract

AMPS in relation to consequences of testing and test fairness was a focus of all studies (Studies I to IV).

Results

In Study I, data for 1346 participants from Middle Europe and 144,143 participants from other world regions were analyzed. The participants were between the ages of 3 and 103 years, and they were well or had a variety of diagnoses. The results revealed that overall the item difficulty calibration values of the AMPS remained stable and that only one out of 36 ADL items of the AMPS demonstrated DIF, but this DIF did not lead to DTF (i.e., all measures fell within 95% confidence bands).

In Study II, data for 11,189 typically-developing children from Middle Europe and other world regions who were between the ages of 2 and 15 were analyzed. The results of ANOVAs revealed significant effects for mean ADL motor and for ADL process ability measures by region and a significant age by region interaction effect for mean ADL process ability. Out of 168 estimated contrasts between Middle Europe and the other world regions for mean ADL motor and ADL process ability, only seven were statistically significant (4.17%), and only two were more than ±1 SE from the international means.

In Study III, regression analyses of data for 10,998 children, 4 to 15 years, who were typically-developing or with mild disabilities, revealed significant age by group interaction effects. Post hoc t tests revealed significant group differences in ADL ability at all ages beyond the age of 4. ADL process ability effect sizes were moderate to large at all ages and ADL motor ability effect sizes were mostly moderate to large age 6 and above.

In Study IV, the use of the AMPS within the context of a feasibility study based on data of 17 Swiss children with mild disabilities was evaluated. The analyses revealed several strengths and problems that were related to the time, equipment, and materials for administering the AMPS, the adherence to standardized administration procedures, the scope of the AMPS as a test of ADL performance, and the reliable rating by the blinded rater.

Conclusion

This thesis provided evidence to support the validity of the AMPS measures and scales when used to evaluate quality of ADL task performance of persons from Middle Europe. Additionally, this thesis provided evidence that the international age-normative means of the AMPS are likely applicable to children from Middle Europe. Moreover, the findings supported the sensitivity of the AMPS measures to discriminate between typically-developing children and children with and at risk for mild disabilities. When it comes to implementation of the AMPS in the context of a feasibility study, the findings indicated both strengths and problems in using the AMPS as an outcome measure that need to be considered when planning further studies.
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADHD</td>
<td>Attention deficit/hyperactivity disorder</td>
</tr>
<tr>
<td>ADL</td>
<td>Activities of daily living</td>
</tr>
<tr>
<td>AIC</td>
<td>Akaike information criterion</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>AMPS</td>
<td>Assessment of Motor and Process Skills</td>
</tr>
<tr>
<td>BI</td>
<td>Barthel Index</td>
</tr>
<tr>
<td>B</td>
<td>Unstandardized regression coefficient</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>CO-OP</td>
<td>Cognitive Orientation to Daily Occupational Performance</td>
</tr>
<tr>
<td>COPM</td>
<td>Canadian Occupational Performance Measure</td>
</tr>
<tr>
<td>d</td>
<td>Cohen’s measure of effect size</td>
</tr>
<tr>
<td>DCD</td>
<td>Developmental coordination disorder</td>
</tr>
<tr>
<td>DIF</td>
<td>Differential item functioning</td>
</tr>
<tr>
<td>DTF</td>
<td>Differential test functioning</td>
</tr>
<tr>
<td>EBP</td>
<td>Evidence-based practice</td>
</tr>
<tr>
<td>ES</td>
<td>Effect size</td>
</tr>
<tr>
<td>FIM</td>
<td>Functional Independence Measure</td>
</tr>
<tr>
<td>g</td>
<td>Hedges’ measure of effect size</td>
</tr>
<tr>
<td>ICF</td>
<td>International Classification of Functioning, Disability and Health</td>
</tr>
<tr>
<td>IADL</td>
<td>Instrumental activities of daily living</td>
</tr>
<tr>
<td>LD</td>
<td>Specific learning disability</td>
</tr>
<tr>
<td>M</td>
<td>Sample mean, arithmetic average</td>
</tr>
<tr>
<td>Mdn</td>
<td>Median</td>
</tr>
<tr>
<td>ME</td>
<td>Middle Europe</td>
</tr>
<tr>
<td>MFR</td>
<td>Many-facet Rasch</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean square error</td>
</tr>
<tr>
<td>MS</td>
<td>Mean square</td>
</tr>
<tr>
<td>n</td>
<td>Number of cases</td>
</tr>
<tr>
<td>OTIPM</td>
<td>Occupational Therapy Intervention Process Model</td>
</tr>
<tr>
<td>PADL</td>
<td>Personal activities of daily living</td>
</tr>
<tr>
<td>PEDI</td>
<td>Pediatric Evaluation Disability Inventory</td>
</tr>
<tr>
<td>PhD</td>
<td>Doctor of Philosophy</td>
</tr>
<tr>
<td>PRPP</td>
<td>Perceive Recall Plan Perform System of Task Analysis</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomized Controlled Trial</td>
</tr>
<tr>
<td>$R^2$</td>
<td>Measure of strength of association</td>
</tr>
<tr>
<td>RSS</td>
<td>Residual sum of squares</td>
</tr>
<tr>
<td>$SD$</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>$SE$</td>
<td>Standard error</td>
</tr>
<tr>
<td>$z$</td>
<td>A standardized score</td>
</tr>
<tr>
<td>ZHAW</td>
<td>Zurich University of Applied Sciences</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Probability of making a type I error</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Probability of making a type II error</td>
</tr>
<tr>
<td>$\eta^2$</td>
<td>Eta squared, measure of strength of relationship</td>
</tr>
</tbody>
</table>
ORIGINAL PAPERS


Original papers have been reproduced with kind permission from the publishers.
Preface

I used to work in Austria and Switzerland as an occupational therapy practitioner. My experience was mainly in the areas of neurological rehabilitation and pediatrics. In 1998, when I started to work as an occupational therapist, it was quite easy to get funding for occupational therapy for clients who were in need of occupational therapy services. However, this situation has rapidly changed due to expanding costs to the insurance system. Nowadays, physicians and occupational therapists are often denied funding for occupational therapy services, especially in the field of pediatrics, and they are constantly asked to provide evidence for the effectiveness of the implemented services.

“How can we evaluate the effectiveness of our occupational therapy services?” was a challenging question that arose when I worked in clinical practice. This question has kept me busy for many years. You may wonder what exactly was the challenge. To me, the biggest one was the lack of valid evaluation tools to evaluate the effectiveness of occupational therapy interventions focusing on clients’ performances of daily life tasks and their participation in different areas of life. It was not a problem at all to evaluate and measure a client’s resources and impairments in terms of body functions, such as range of motion, because evaluation tools to test these were available, even in Middle Europe. No, it was a problem to find valid outcome measures at the level of activity and participation.

During the years that I worked in pediatric practice, I met children who were referred to occupational therapy because they had problems in performing daily life tasks at school and at home. These problems were often related to ordinary activities like putting on and tying shoes, manipulating zippers and buttons, or preparing a sandwich. At the first sight, these problems did not appear as serious as were their academic problems, but I realized that they had a serious impact on the children’s self-esteem; on their family lives; their relationships with parents, siblings, and peers; and their participation in community life. What does it mean, for example, for a 7-year-old if he or she needs support from his teacher to open and close his or her trousers when he or she goes to the toilet? Might he or she avoid asking for support or feel ashamed when

---

1 In the introduction of this thesis, I have used the first person singular (I, my) as it originated from my experience as an occupational therapy practitioner. In the other parts of the thesis, I use the first person plural (we, our) in order to show that my research has always been implemented in collaboration with my co-authors.

2 In this thesis I use the term evaluation tool to specify a broad range of devices that are in the literature and variously referred to as tests, assessments, scales, inventories, and/or measures.
doing so? However, when working with children, one focus was to enable children’s performances of daily tasks and support parents and teachers to find long-term strategies and solutions; the other was to convince insurance companies about the benefits of occupational therapy interventions in order to receive funding.

Therefore, I was in need of an evaluation tool that could be used to evaluate children’s strengths and limitations in occupational performance and to evaluate and document the effectiveness of my services. I started to search for standardized evaluation tools, was trained in some, and implemented them in my daily clinical practice. I found them helpful in many ways—for evaluating clients’ actual performances, setting goals, planning targeted interventions, evaluating outcomes, and communicating with clients, medical doctors, and insurance companies. Yet, none of these evaluation tools had been validated for use in Middle Europe or with children. What a disappointment. What an inspiration for a dissertation.

I took up the idea and now I have finished. You might get the impression that I focused on measuring and numbers—no—I am just convinced that the use of valid outcome measures is essential for occupational therapy practice. Therefore, I hoped to contribute to the field of occupational therapy in general and in Middle Europe specifically—to provide further evidence for an occupation-based and occupation-focused evaluation tool. Nevertheless, my work is a drop that may blend in the ocean, but I hope that there will be a ripple effect and that my contributions will be continued by occupational therapists in practice and research in order to provide continued evidence for the effectiveness of evaluation and intervention services and ensure occupational therapy for our clients in the long-term.
INTRODUCTION

Occupation is the core of occupational therapy (Kielhofner, 2008; Townsend & Polatajko, 2007) and is used as the means and goal of occupational therapy interventions (American Occupational Therapy Association [AOTA], 2008; Latham, 2008). So, occupational therapists also use occupation-based and occupation-focused methods for evaluation and/or intervention (Fisher, 2013). More specifically, by engaging persons in occupations as their main method, they work occupation-based (e.g., they evaluate a person’s strengths and weaknesses by observing how he or she performs daily life tasks), and by putting the immediate focus on occupation, they work occupation-focused (e.g., they evaluate the person’s perceived strengths and weaknesses by asking how he or she experiences daily life task performances) (Fisher, 2013). Thus, occupational therapists evaluate and promote person’s abilities to engage in daily life tasks that they are expected, need, and/or want to perform (Fisher, 2009).

The purpose of this thesis was to contribute evidence to support the use of an occupational-based and occupational-focused evaluation tool—the Assessment of Motor and Process Skills (AMPS) (Fisher & Jones, 2012a, 2012b). More specifically, I aimed to evaluate validity evidence from different sources of the AMPS scales and measures that support its use with children internationally, as well as its use as an occupation-based and occupation-focused evaluation tool in Middle Europe, including its use to help guide occupation-based and occupation-focused interventions. In the following sections I define the concepts of children with mild disabilities and children at risk as I used them in this thesis, describe their problems in performing activities of daily living (ADL) tasks, and summarize effectiveness evidence for occupational therapy used for these children. Then, I describe occupational therapy in the context of Middle Europe with a focus on the use of valid occupation-based and/or occupation-focused evaluation and intervention methods. Furthermore, I specify the concept of occupational performance and how it is evaluated in the occupational therapy process. I then describe the AMPS scales and measures. Finally, I describe the concept of validity in general, in relation the AMPS specifically, and scientific approaches used for evaluating validity evidence from different sources. After the introduction, I specify the aims of the thesis, describe the methods and results, discuss the results, and finish with a conclusion.
Occupational Therapy With Children and in the Context of Middle Europe

Children With and at Risk for Mild Disabilities

In this thesis, children with mild disabilities are children with developmental coordination disorder (DCD), disorders of attention (e.g., attention-deficit/hyperactivity disorder, ADHD), specific learning disabilities (LD) (e.g., speech and language disorders, specific reading disabilities), and sensory integrative dysfunction (SI). Children with ADHD, DCD, and LD account for 5% to 15% of the population (American Psychiatric Association [APA], 2013). The diagnoses often co-exist (APA, 2013; Dewey, Crawford, Wilson, & Kaplan, 2004; Dewey, Kaplan, Crawford, & Wilson, 2002) and symptoms are often overlapping. For example, Kadesjö and Gillberg (1998) found that approximately half of the children diagnosed with DCD also had moderate to severe symptoms of ADHD. Therefore, Gillberg (2010) proposed to unite these children with diverse symptoms, between the ages of 3 and 5, under the term ESSENCE (Early Symptomatic Syndromes Eliciting Neurodevelopmental Clinical Examinations) rather than split them into diverse diagnostic groups. Their problems typically persist through adolescence and adulthood (APA, 2013; 2000; Kadesjö, 2001; Kadesjo & Gillberg, 1998; Shaw et al., 2012), meaning that the impact on individuals and society as a whole is enormous in terms of financial costs, stress to families, and the impact on schools (APA, 2013; Weller & McDermott, 2007).

In this thesis, I refer to children with DCD, ADHD, LD, and SI as children with mild disabilities. I claim that from a client-centered, occupation-based, and occupation-focused perspective, children with these diagnoses can be united. My arguments are multifaceted, as I describe in the following:

To start with, the term disability has traditionally been viewed in terms of individuals having impairments of physical (e.g., neuro-musculoskeletal) and mental functioning (e.g., attention) that restricted their participation in mainstream activities (Priestly, 2003). In recent years, the focus has shifted from this medical view towards a more social view of disability which highlights that disability is not caused by impairment per se, but is a product of complex social structures and processes (Barnes, Mercer, & Shakespeare, 1999; Priestly, 2003). This social view of disability better fits contemporary occupational therapy than does a medical-centered one. Contemporary occupational therapy highlights the interrelationship between the client, the environment, the daily tasks performed, and occupational performance (Fisher, 2009; Kielhofner, 2008; Townsend & Polatajko, 2007).

Furthermore, children with ADHD, DCD, LD, and SI have in common that their primary impairments significantly interfere with ADL and
Introduction

school task performance (APA, 2013; Barkley & Fischer, 2011; Munkholm & Fisher, 2008). Therefore, interpreted from an occupation-focused perspective, these children share restrictions performing and participating in daily activities in different environments.

Moreover, compared to children with severe disabilities, the problems of children with mild disabilities are often not readily visible when they are observed performing daily life tasks in natural environments (Pellegrino, 2007). This phenomenon is supported when standardized evaluation tools like the Movement ABC (Henderson, Sudgen, & Barnett, 2007) or the AMPS (Fisher & Jones, 2012a, 2012b) are used to evaluate for differences between groups. That is, the results of different studies showed that children with mild disabilities have in common that their test performance is often between 1 and 2 standard deviation (SD) below the mean (Geuze, Jongmans, Schoemaker, & Smits-Engelsman, 2001), whereas the test performance of children with severe disabilities is often more than 2 SD below the mean. When using the AMPS, results of a recent study revealed that ADL task performance of children with spina bifida was often 2 SD below the mean (e.g., Peny-Dahlstrand, Åhlander, Krumlinde-Sundholm, & Gosman-Hedström, 2009).

To conclude, there are several arguments that support the use of the term mild disabilities in a thesis that focuses on client-centered, occupation-based, and occupation-focused evaluation and intervention. However, my conclusion might be rebutted by the argument that occupational therapy, as a health profession, has roots in the medical field (Meyer, 1922) and is, at least in Middle Europe, dependent upon the referral of medical doctors (Bundeskanzleramt Österreich, 2013; Bundesministerium für Justiz, 2011; Bundesversammlung der Schweizerischen Eidgenossenschaft, 2006; Ministrrtvo za Zdravstvo, 2005). Therefore, the argument can be proposed that we should use diagnostic medical terms. Nevertheless, I assert that a profession that has its focus on occupation should use terms that reflects its focus. Using the term mild disabilities to unite these children is nothing new and is commonly done in educational research (e.g., Cook, 2001) and has recently been used in occupational therapy research (Munkholm, Berg, Lofgren, & Fisher, 2010). However, a thorough argument for using this term in occupational therapy research has not been published before.

Another concept that I use in this thesis is the concept of children at risk. I use the term children at risk to refer those children who are not diagnosed, but are experiencing problems and are therefore at risk of developing and being diagnosed with a disability (Heward, 2014).

The concept of children at risk is commonly used in educational (Heward, 2014; Johnson, 1998; Schoemaker, Flapper, Reinders-Messelink, & de Kloet, 2008; Steele, 2004), psychological (Coleman, Piek, & Livesey, 2001), medical (Coghill & Sonuga-Barke, 2012), and sociological (Lubeck & Garrett, 1990) research and often refers to
children who experience problems in school and life because they live in poverty, or are neglected or abused (Heward, 2014; Johnson, 1998; Lubeck & Garrett, 1990). In this thesis, I use the term in a slightly different way, highlighting that these children do have observable problems and/or experience problems (Heward, 2014). The term children at risk has been currently implemented and used in relation to occupational therapy evaluation tools (Fisher, Bryze, Hume, & Griswold, 2007; Fisher & Griswold, 2010; Fisher & Jones, 2012a, 2012b) and occupational therapy research (Munkholm et al., 2010; Munkholm & Fisher, 2008). Like the concept of mild disabilities, the concept of being at risk can be interpreted from both an individual or a societal perspective, where performance problems are grounded in the individual alone or the individual’s interaction with the environment (Johnson, 1998; Lubeck & Garrett, 1990). Occupational therapy theory highlights the interaction of the individual with the environment, meaning that performance and performance problems are always influenced by the person, the specific task, the environment, and the society and culture (Fisher, 2009; Law & Dunbar, 2007; Loukas & Dunn, 2010; Palisano et al., 2012).

ADL Problems of Children With Mild Disabilities

Children with mild disabilities might have different medical diagnoses, but they share problems in performing daily life tasks (e.g., ADL and school task performance) (APA, 2013). The results of earlier studies provide evidence of experienced and observed problems in ADL task performance of children, which are summarized below.

In 2011, a systematic review by Magalhães, Cardoso, and Missiuna (2011) showed that the most frequent activity limitations and participation restrictions of children with DCD were in the areas of play, classroom activities, sports, ADL, and social skills. When ADL was considered, they found that 47.72% of the reviewed studies reported that children with DCD had limited abilities to perform self-care tasks (e.g., dressing, using cutlery, feeding independently). No systematic reviews or meta-analyses were found of children with ADHD, SI, or LD, but many other studies investigated ADL task performance problems of the group of children with mild disabilities. For example, earlier studies found that children with mild disabilities experience difficulties in performing self-care tasks, and coping with time pressures to get ready (Dunford, Missiuna, Street, & Sibert, 2005; Segal, 1998; Summers, Larkin, & Dewey, 2008); they had significantly lower ADL task performance than did typically-developing children (Gol, Jarus, Gol, & Jarus, 2005; Rodger et al., 2003; Wang, Tseng, Wilson, & Hu, 2009; White & Mulligan, 2005; White, Mulligan, Merrill, & Wright, 2007). One study compared household task performance of typically-developing children and children with ADHD and found no significant differences in frequency of performance, but significant differences in the level of assistance needed (Dunn, Coster,
Orsmond, & Cohn, 2009). Considered together, all these studies have shown that children with mild disabilities are at greater risk for limited ADL task performance.

Moreover, because they lack the skills needed to participate fully in society, it is common that children with mild disabilities experience failure, come to expect further failure, and develop lower self-esteem that affects their social, academic, and physical performance. As a result, they are more likely to demonstrate alcohol abuse, increased criminality, reading disorders, lower educational level, and low self-esteem as adolescents and adults (Rasmussen & Gillberg, 2000; Shaw et al., 2012). Therefore, early identification and therapeutic intervention is particularly important for these children, their parents, and society as a whole (Gillberg, 2010). Shaw et al. (2012) conducted a systematic review of long-term outcomes (i.e., academic, antisocial behavior, driving, non-medicinal drug use/addictive behavior, obesity, occupation, services use, self-esteem, and social function) in persons with ADHD and found that persons with ADHD, when left untreated, had lower outcomes compared to persons with ADHD who received non-pharmaceutical and pharmaceutical interventions or to persons without ADHD.

Effectiveness of Occupational Therapy With Children With Mild Disabilities

With their expertise in enabling occupation, occupational therapists can provide a multitude of successful intervention strategies for children with mild disabilities (e.g., self-management techniques, social skills training, environmental adaptations, parent education) for enabling performance of ADLs and participation (Rodgers, 2005). Despite this knowledge, only a few occupational-therapy-relevant research studies have been implemented (Bennett et al., 2007). To start with occupation-based and/or occupation-focused approaches, several studies were conducted to evaluate CO-OP (Cognitive Orientation to Daily Occupational Performance) intervention in children with DCD and showed that children with DCD could develop effective strategies to perform ADLs (Taylor, Fayed, & Mandich, 2007; Ward & Rodger, 2004). Furthermore, a pilot randomized controlled trial (RCT) in children with DCD showed that CO-OP was more effective than a treatment approach that focused on the restoration of body functions (Miller, Polatajko, Missiuna, Mandich, & McNab, 2001).

Other studies have evaluated the effectiveness of occupational therapy approaches for developing body-functions and showed that these approaches resulted in improved motor performance in children with DCD (Davidson & Williams, 2000; Leemrijse, Meijer, Vermeer, Ader, & Diemel, 2000). Another study showed that children with ADHD could significantly improve both occupational performance of daily life tasks (e.g., keeping room tidy, riding a bike) and executive functioning...
Introduction

following a cognitive-functional-focused occupational therapy intervention program (Hahn-Markowitz, Manor, & Maeir, 2011).

Chu and Reynolds (2007) evaluated the effectiveness of a combined occupational therapy approach in children with ADHD. The intervention included education of parents and teachers, behavioral management, a diet program, adaptation of the environment, and restorative and aquisitional occupational therapy interventions. They found that 13 out of 20 children with ADHD improved significantly in their behavior due to the combined occupational therapy intervention. They highlighted the role of the parents in the intervention process and suggested further investigation of combined occupational therapy approaches. Another study evaluated the effectiveness of a play-based occupational therapy intervention for children with and without ADHD that combined the use of play, feedback, and parental education and found a large effect in improving the social play in all children (Wilkes, Cordier, Bundy, Docking, & Munro, 2011).

Finally, one systematic review and one meta-analysis of the efficacy of occupational and physical therapy interventions with children with DCD have been published. They showed strong evidence that intervention, in general (e.g., any kind of occupational therapy), is better than no intervention (Hillier, 2007) and specifically, that task-oriented approaches (e.g., CO-OP) were more effective than were approaches that focused on development and restoration of body functions (e.g., sensory integrative approach) (Hillier, 2007; Pless & Carlsson, 2000). The authors clearly recommended implementing task-oriented approaches over the age of 5. Findings with regard to intervention duration were not clear (Pless & Carlsson, 2000). No systematic reviews or meta-analyses were published related to children with ADHD, LD, and SI; not about occupational therapy interventions in general or their ADL task performance problems specifically.

To conclude, the existing research only supports the effectiveness of occupational therapy intervention in children with ADHD and DCD. All trials were rather small, with single-case or before-after design and only one of 12 studies used an RCT design (Miller et al., 2001). Therefore, additional intervention studies that support the effectiveness of occupational-based and occupation-focused interventions with children with mild disabilities are critically needed.
Occupational Therapy in the Context of Middle Europe

It is critical to place this project into the context of current occupational therapy practice in Middle Europe. Traditionally, occupational therapy evaluations and interventions in Middle Europe have mainly focused on restoration of body functions and body structures. For example, the use of the sensory integrative approach (Bundy, Lane, & Murray, 2007) has been commonly used and many occupational therapists continue to formulate therapeutic goals at the impairment level rather than at the level of occupation (i.e., activity and participation level) (Costa & Schönthaler, 2011; Romein, 2004). This means that the implementation of occupation-based and/or occupations-focused evaluations and interventions has been less common, but has been increasing steadily.

While the use of standardized occupational-based and occupational-focused evaluation tools is slowly increasing, not many among those used have been validated for use with Middle Europeans. Only, the Canadian Occupational Performance Measure (COPM) (Law et al., 2009)–an occupation-focused evaluation tool—is well known and widely used. The COPM is very valuable for enabling clients to prioritize their concerns and establish client-centered goals. Otherwise, there is a critical gap in valid occupation-based and/or occupation-focused evaluation tools (Projektgruppe ergotherapeutische Befundinstrumente in der Pädiatrie, 2004; Röse & Seitz, 2008).

This situation has occurred for several reasons. First, while Middle European countries share values and beliefs based on an evocative history, they use a variety of languages (i.e., German, French, Italian, Retro-Roman, and Slovenian) and English is not ranked among them (Raunig, 2010). Occupational therapy evaluation tools that have been developed in English-speaking countries are often not available in the languages spoken and/or are not validated for the use in Middle Europe. Second, only a few evaluation tools have been developed within Middle Europe and many of them were developed without establishing their validity and reliability (Jobst, 2008; Siebert, Süess-Marz, Henkes, & Köster, 2007). Third, the majority of evaluation tools used in Middle European settings are body-function-focused (Siebert et al., 2007) rather than occupation-based and/or occupation-focused. For example a book that summarizes evaluation tools in the field of neurological rehabilitation in Middle Europe, covers 82 evaluation tools, but only six focus on evaluating independence in ADLs (Schädler et al., 2011). Evaluation tools that can be used to evaluate a person’s performance in other activity and participation areas beyond ADL and mobility are missing (Schädler et al., 2011). Fässler and Marchner (2013) implemented a survey of Swiss neurological rehabilitation centers and found that the most common evaluation tools used to evaluate ADL were
Introduction

the Functional Independence Measure (FIM) (Granger, Greer, Liset, Coulombe, & O’Brien, 1975), the Perceive Recall Plan Perform (PRPP) System of Task Analysis (Chapparo & Ranka, 2003), the AMPS (Fisher & Jones, 2012a, 2012b), and the Barthel Index (BI) (Fässler & Marchner, 2013; Mahoney & Barthel, 1965). The FIM and the BI are ADL evaluation tools for evaluating independence, safety and frequency of assistance in basic ADLs and have partly been validated for use in Middle Europe (e.g. Heuschmann et al., 2005; Lundgren-Nilsson et al., 2005). The PRPP is an occupation-based evaluation tool that has its primary focus on cognition (i.e., is not occupation-focused), and the AMPS is an occupation-based and occupation-focused evaluation tool used to evaluate a person’s quality of ADL task performance. For the field of pediatrics, no ADL evaluation tool has been validated for use in Middle Europe. However, a project to culturally adapt and validate the Pediatric Evaluation of Disability Inventory (PEDI) (Haley, Coster, Ludlow, Haltiwanger, & Andrellos, 1992) for German-speaking countries is on-going (Schulze, Page, Kottorp, & Lilja, 2013).

Obviously, there is a need to develop knowledge in the area of occupation-based and/or occupation-focused evaluations. Therefore, the evaluation of validity evidence from different sources of the AMPS for use in Middle Europe is the focus of my thesis. One reason for choosing the AMPS evolved from my occupational therapy practice. I found the AMPS useful as it can be used with clients of different ages (children, youth, and adults) and with a variety of diagnostic groups to evaluate their strengths and limitations in ADL task performance. Furthermore, I found the AMPS reports very helpful for communication with clients, physicians, and insurance companies. Moreover, I could easily use the AMPS to evaluate and document the effectiveness of my occupational therapy services on an individual level. Another reason for choosing the AMPS was that it was already a well-established evaluation tool used to test quality of ADL task performance, which had already been used in many world regions and settings, and with a large base of evidence (Fisher & Merritt, 2012b). Last, but not least, the number of occupational therapists in Middle Europe who have successfully completed training in the administration and interpretation of the AMPS, and who have been calibrated a reliable and valid AMPS raters, has been increasing. But even though the AMPS has been developed as an international, cross-regional ADL evaluation tool, it was not validated for use in Middle Europe.
Evaluating ADL Task Performance

Concept of Occupational Performance

Occupational performance is a central concept in occupational therapy that is described in many models (Fisher, 2009; Kielhofner, 2008; Nieuwesteeg-Gutzwiller & Somazii, 2010; Polatajko, Townsend, & Craik, 2007) and refers to the act of doing (Baum & Christansen, 2005; Nelson, 1988, 1997). For example, a child who is brushing his or her teeth is performing a daily life task he or she needs and is expected to perform. Occupational performance is influenced by every day habits, routines, and roles (Kielhofner, 2008; Kramer & Bowyer, 2007; Law & Dunbar, 2007). More specifically, brushing one’s teeth is influenced by the patterns of action a child uses when brushing his or her teeth (habits), when and where he or she brushes his or her teeth (routine), and by being a son or daughter who is expected to perform brushing one’s teeth when he or she is asked to do so (role).

More specifically, occupational performance is closely related to desired roles, defined as positions in society, each having a defined status and specific expectations for behavior (e.g., Baum & Christansen, 2005; Chard, 2010; Law & Dunbar, 2007). Roles are dynamic and change with development and skill acquisition over time. Roles of children include, for example, the role of a son or daughter, friend, self-carer, school student, and team member (Rodger & Ziviani, 2006). These roles shape what children do and how they view the world. They also profoundly influence who children are, that is, their identities (McColl, 2010). Moreover, meaningful engagement in desired roles both allows for and is dependent upon a child’s participation in the daily task performances (i.e., occupations) associated with their desired roles in their homes, communities, and schools (Kramer & Bowyer, 2007). Occupational therapists, therefore, promote skill acquisition and enable children to develop competence in their roles. For example, through engagement in ADLs, children are enabled to competently perform daily life tasks needed to participate in their homes, schools, and community contexts (Loukas & Dunn, 2010). One specific focus of this thesis is on the child’s ability to perform personal ADL (PADL, self-care) and instrumental ADL (IADL, e.g., household chores) tasks at home or school.

Additionally, the environment, society, and culture impact a person’s performance (Fisher, 2009). It is generally agreed that occupational performance is the output of the relationship between the person, the daily life task, and the environment (e.g., Fisher, 2009; Polatajko, Davis, Cantin, Dubouloz-Wilner, & Trentham, 2007). Persons perform daily life tasks in different areas of life, such as PADL and IADL, work, and leisure (AOTA, 2008). ADL task performance—the focus of this thesis—is one area of occupational performance.
The concept of occupational performance is closely related to the concept of participation as described by the International Classification of Functioning, Disability and Health (ICF) (World Health Organisation [WHO], 2001, 2007). In the ICF, participation is viewed as functioning in a natural environment or person-specific context and defined as “involvement” in a life situation (WHO, 2001, p. 17). Both concepts—participation and occupational performance—have in common that they cover performance of daily live tasks. However, they differ in one important aspect: the concept of participation does not cover the subjective experience of doing whereas the concept of occupational performance does (Gantschnig, Hemmingsson, & La Cour, 2011; Hemmingsson & Jonsson, 2005; Pereira, la Cour, Jonsson, & Hemmingsson, 2010; Polatajko, Davis, Cantin, et al., 2007). Occupational therapists believe that it is never enough to simply observe how a person performs a daily life task. They also emphasize the person’s meaning and satisfaction with performance (Fisher, 2009; Polatajko, Davis, Cantin, et al., 2007).

Evaluating ADL Task Performance Within the Occupational Therapy Process

Occupational therapists support persons in engagement of meaningful and purposeful daily life tasks (Fisher, 2009; Kielhofner, 2008; Polatajko, Townsend, et al., 2007) in different performance areas and, in this way, promote their health and well-being (Townsend et al., 2007). When persons seek occupational therapy services, the process usually starts with exchanging information related to the person’s daily occupations in his or her context, and his or her experienced strengths, problems, and priorities of occupational performance (Chard, 2010; Fisher, 2009; Nelson & Jepson-Thomas, 2003). Then, nonverbal information is gathered by implementing a performance analysis through observing person’s quality of performance of daily life tasks (AOTA, 2008; Fisher, 2009; Fisher & Jones, 2012a).

An occupational performance analysis is an important part of the evaluation process (AOTA, 2008; Fisher, 2009). Occupational therapists implement performance analyses to evaluate quality and the effectiveness of a person’s task performance in a natural environment (Chard, 2010; Fisher, 2009; Fisher & Jones, 2012a). More specifically, they evaluate the person’s task performance in terms of the smallest observable units of occupational performance, called performance skills, which can be group under motor, process, and communication and social (interaction) skills (AOTA 2002; 2008; Fisher, 2009). For example, when an occupational therapist observes how a person performs everyday tasks, she or he gathers information about the skills that are performed effectively and support task performance (e.g., competently grasp the tooth brush) and those that are performed ineffectively and delay or disrupt task
performance (e.g., clumsiness when manipulating the toothpaste tube in-hand).

Occupational performance skills are not underlying body functions, they are small actions of occupational performance (Fisher, 2009; Fisher & Jones, 2012a). Capacities and impairments of body functions and body structures, resources and barriers in the environment, and personal factors are only considered after a performance analysis has been completed (Chard, 2010; Fisher, 2009; Fisher & Jones, 2012a).

In order to evaluate occupational performance of persons and plan effective occupational therapy interventions, it is important to use reliable, valid, and sensitive evaluation tools that are client-centered, occupation-based, and/or occupation-focused (Law et al., 2009; Law, Baum, & Dunn, 2005; Polatajko, Davis, Stewart, et al., 2007) and compatible with the ICF (Law et al., 2005; World Health Organization, 2001, 2007). The AMPS is one ranked among these.

Evaluating ADL Task Performance Using the AMPS

The specific focus of this thesis was evaluating validity evidence from different sources of the AMPS (Fisher & Jones, 2012a, 2012b). The AMPS was designed to be used by occupational therapists to measure the quality of a person’s performance of ADL tasks in naturalistic settings. Thus, the AMPS is occupation-based and occupation-focused. Furthermore, the AMPS is a criterion-referenced evaluation tool that also allows for norm-referenced interpretations. The specific characteristic of a criterion-referenced evaluation tool is that a person’s performance is evaluated against the defined criterion of competence; whereas, the specific characteristic of a norm-referenced evaluation tool is that person’s performance is compared to that of other persons of the same age (Anastasi & Urbina, 1997; Browning, 1997; Fawcett, 2007). More specifically, in the AMPS, a person’s ADL task performance is primarily compared to the criterion of competent quality of ADL motor and ADL process item performance (e.g., the occupational therapist observes and judges the person to “readily and consistently raise or lift task objects, including lifting an object from one place to another, but without ambulating or moving from one place to another.”) (Fisher & Jones, 2012b, p. 8-14). Additionally, the AMPS software can be used to generate reports that include age norms that can become another important point of reference from which to judge whether or not a person’s ADL task performance is of concern, suggesting an indication for a person’s need of occupational therapy intervention.

Currently, there are more than 120 standardized ADL tasks in the AMPS that are hierarchically ordered according to their task challenges (see Figure 1). While some ADL tasks are generally viewed as being world-
Introduction

region-specific (e.g., *Peanut butter and jelly sandwich* for use in North America, *Eating an Asian meal with chopsticks* for use in Asian countries), the majority of the AMPS tasks are among the most commonly performed ADL tasks internationally (e.g., *Upper and lower body dressing, Cleaning a bathroom*).

The standardized administration procedures of the AMPS include several steps. They start with interviewing the person, where he or she is to choose two ADL tasks from among a subset of the 120 ADL tasks included in The AMPS manual based on the following criteria: a) they are meaningful for and relevant to the person’s daily life, b) they are ADL tasks that currently are presenting a challenge, and c) the person has prioritized them for further evaluation and intervention (Fisher & Jones, 2012a). Before the ADL tasks are performed, the person and the therapist explore and become familiarized with the environment. Then, no matter which two tasks the person is observed performing, he or she is scored on the same 16 ADL motor and 20 ADL process items, once for each task performed. The occupational therapist rates the person’s performance on each of the 36 ADL items using a four-point ordinal scale (see Table 1).

<table>
<thead>
<tr>
<th>Easier ADL tasks</th>
<th>Easier ADL motor items</th>
<th>Easier ADL process items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eating a snack with an utensil</td>
<td>Lifts</td>
<td>Uses</td>
</tr>
<tr>
<td>Brushing teeth</td>
<td>Moves</td>
<td>Sequences</td>
</tr>
<tr>
<td>Folding a basket of laundry</td>
<td>Transports</td>
<td>Searches/Locates</td>
</tr>
<tr>
<td>Feeding a cat — dry cat food and water</td>
<td>Grips</td>
<td>Gathers</td>
</tr>
<tr>
<td>Showering</td>
<td>Bends</td>
<td>Terminates</td>
</tr>
<tr>
<td>Hot cereal &amp; beverage</td>
<td>Coordinates</td>
<td>Restores</td>
</tr>
<tr>
<td>Vacuuming the inside of an automobile</td>
<td>Calibrates</td>
<td>Notices/Responds</td>
</tr>
<tr>
<td>Vegetable soup, vegetables sanitized</td>
<td>Positions</td>
<td>Accommodates</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Harder ADL tasks</th>
<th>Harder ADL motor items</th>
<th>Harder ADL process items</th>
</tr>
</thead>
</table>

*Figure 1. Selected ADL tasks and ADL motor and ADL process items included in the AMPS, arranged from easier to perform to harder to perform (from Gantschnig, Page, & Fisher, 2012, p. 153).*
Introduction

The ADL motor and ADL process items can be thought as a common version of the AMPS, whereas the ADL tasks performed are unique (see Figure 1). This means that if a person is observed brushing his or her teeth, he or she is scored on how effective he or she was as he or she lifted the toothbrush and moved the toothbrush back and forth across teeth. If the person is then observed folding a basket of laundry, he or she is scored, for example, on how effective he or she was as he or she lifted the shirt and moved the hand along the shirt to smooth the wrinkles. As the ADL tasks become harder, each of the ADL skills (e.g., Lifts and Moves) also become proportionally more difficult (see Figure 1). In this way, the AMPS can be thought of being comprised of two item banks, one for ADL motor and one for ADL process items. Within the current item bank, each ADL item is represented 120 times, once for each ADL task included in the AMPS manual (Fisher, 1993; Fisher & Jones, 2012a). After the observation of the person’s task performances, the occupational therapist enters the person’s raw scores into the AMPS software that is used to convert these ordinal data into linear ADL motor and ADL process ability measures, expressed in logits. The AMPS software can then be used to generate AMPS Graphic and Summary Reports (AMPS Project International, 2010). An ADL ability measure above the criterion-referenced cutoff measures of 2.0 logits on the ADL motor scale and 1.0 logit on the ADL process scale most commonly indicate a physical effort, and an efficient, safe, and independent ADL task performance (Fisher & Jones, 2012a). The person’s ADL motor and ADL process ability measures are also transformed into standardized z scores, normalized

Table 1
ADL Skill Item Rating Scale Criteria

<table>
<thead>
<tr>
<th>Score</th>
<th>Quality of performance</th>
<th>Outcome yielded</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Competent</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No evidence of problem</td>
</tr>
<tr>
<td>3</td>
<td>Questionable</td>
<td>Uncertain</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Possible problem</td>
</tr>
<tr>
<td>2</td>
<td>Ineffective</td>
<td>Undesirable use of time</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Undesirable amount of physical effort</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Possible risk</td>
</tr>
<tr>
<td>1</td>
<td>Severely deficient</td>
<td>Unacceptable use of time</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unacceptable amount of physical effort</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Imminent safety risk</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Task breakdown</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Need for assistance</td>
</tr>
</tbody>
</table>

*Note.* More detailed criteria for scoring each ADL motor and ADL process item are published in the AMPS manual (From Fisher & Jones, 2012, p. 8-8).
standard scores, and percentile ranks that allow for a norm-referenced interpretation of the AMPS results. Changes in ADL motor or ADL process ability measures of 0.3 logit are considered to be clinically meaningful; changes of 0.5 logit ($2 \, \text{SE}$) for the ADL motor scale and 0.4 logit ($2 \, \text{SE}$) for the ADL process scale are considered to be statistically significant ($p \leq 0.15$) (Fisher & Merritt, 2012b). Based on the results of the AMPS observation and the AMPS reports, the occupational therapist then interprets the results of the AMPS observation for further decision making (e.g., need of intervention, goals of intervention).

The basis of the AMPS software (AMPS Project International, 2010) are many-facets Rasch analyses (MFR) that are used to generate two linear measures of persons abilities, one for the ADL motor and one for the ADL process scale, taking rater severity, task challenge, and item difficulty in account (Linacre, 1987-2012). The AMPS software is a personalized version of an MFR program that includes the occupational therapist’s rater severity calibration value.

Because the AMPS items have been developed to be universal (i.e., observable during any ADL task), and the person only performs ADL tasks that are familiar, culturally-relevant, and chosen, the AMPS measures should remain free of cross-regional bias when used in Middle Europe. Furthermore, the AMPS was designed for use with persons of all ages beyond the age of 2 and can, therefore, be used to evaluate quality of ADL task performance of children and youth (Brown & Chien, 2010; Fisher & Jones, 2012a; Loukas & Dunn, 2010; Richardson, 2010). I will summarize validity evidence from different sources of the AMPS in more detail below.
Validity of Evaluation Tools

Concept of Validity

Validity is an essential feature that guides the development and use of evaluation tools (e.g., tests). Validity has been described by many authors, and the concept has developed over the last century to have a broad scope (Kane, 2006). Nowadays it is generally agreed that validity refers not only to the degree to which a test measures what it is intended to measure (Bortz, Lienert, & Boehnke, 2000; Dawson & Trapp, 2004; Fawcett, 2007; Polit & Beck, 2008), but also to the meaning and interpretation of test scores, and the use and consequences of both testing procedures and test scores (Brennan, 2006; Kane, 2006).

In this thesis, I will use the definition of validity described by the American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) (1999) as “the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (AERA, APA, & NCME, 1999, p. 9). More specifically, when test scores are used and interpreted in relation to different groups of persons (e.g., persons with mild disabilities) or settings (e.g., different world regions), each intended use and the interpretation of the test scores should be validated. According to the AERA et al. (1999), test developers are responsible for furnishing relevant evidence that support the intended use of a test. In contrast, the person who implements the test is responsible for evaluating the evidence that supports the use of that test in the particular setting in which it is to be used (e.g., in pediatric occupational therapy practice in Middle Europe). Therefore, the process of evaluating validity evidence from different sources can be viewed as a shared responsibility of test developers and test users to build arguments to support the use of tests and the interpretation of test scores, including their relevance to the intended use (AERA et al., 1999).

Rather than distinguishing forms of validity (e.g., construct, criterion, content), the AERA et al. (1999) have described validity as a unitary concept where different sources of evidence contribute to the intended interpretation of test scores. Those sources include evidence based on a) test content (e.g., relationship between a test’s content and the construct the test is intended to measure), b) response processes (e.g., fit between the construct and the performance actually engaged in by examinees), c) internal structure (e.g., the degree to which test items conform to the construct), d) relations to other variables (e.g., influence of external variables on validity), and e) consequences of testing (e.g., if there are positive or negative consequences for a specific group when tested). I have used these sources of validity for framing my studies, describing the results, and discussing the results.
Introduction

Approaches for Evaluating Validity Evidence

A cost-effective method for developing standardized evaluation tools for use in a specific world region and with a specific person group is to adapt, as needed, and validate existing tools. There are many different approaches for evaluating validity evidence and I will describe those that are relevant to the use of the AMPS and the interpretation of the AMPS measures for typically-developing children, for children with and at risk for mild disabilities, and for persons from Middle Europe below.

As I described earlier, the ADL motor and ADL process ability measures of the AMPS are generated by MFR analyses embedded in the AMPS software (AMPS Project International, 2010). Specific to Rasch analyses is that person’s raw item scores (i.e., ordinal scores) are converted into a linear measure, expressed in logits. The mean difficulty of the items included in the test is usually set at zero logits. Each person’s ability is then given as the logit difference with respect to this value (Bond & Fox, 2007). Rasch analysis methods also can be used for developing new evaluation tools and validating existing ones. More specifically, Rasch analysis methods can be used to evaluate whether the items of an evaluation tool behave in the same way across groups of persons from different world regions (i.e., if item difficulties remain stable across world regions, without evidence of differential item functioning [DIF]). For example, by comparing the item difficulty calibration values based on a Middle European sample to the item difficulty calibration values based on samples from other world regions, the presence or absence of DIF in Middle Europe can be detected. More specifically, when differences between item difficulty calibration values of one group compared to another display statistical significance and a meaningful DIF contrast, DIF is determined (Draba, 1977; Wilson, 2005). If an evaluation tool is free of DIF among world regions, the item hierarchies will be the same (i.e., stable, independent of region where the evaluation tool is used). In contrast, if meaningful differences in item difficulty calibration values between regions arise, for example if an item is calibrated as easier for persons in Middle Europe than for persons from other world regions, DIF is detected. DIF can mean that persons tested from one world region may be at unfair disadvantage when compared to persons from other world regions. Therefore, it is necessary to determine if DIF leads to differential test functioning (DTF) (Osterlind & Everson, 2009; Pae & Park, 2006).

From a clinical perspective, it is important to evaluate for DTF, because, in clinical practice, decisions (e.g., if a person is in need for occupational therapy services) are based on the person’s performance on the whole test, not on his or her performance on one item. When groups of persons are compared, and their total measures differ, then, DTF is present (Pae & Park, 2006). Thus, cross-regional DTF is usually analyzed by plotting total test measures for all persons when they are estimated based on the item difficulty calibration values from one world region against total test measures for all persons when they are estimated based on the item
difficulty calibrations from another world region. The location of the paired measures (i.e., the total test measures based on one world region paired with the total measures based on another world region) should fall within 95% confidence bands. If measures are located outside the 95% confidence bands, DTF is detected, signaling that test bias is present, and that there is a threat to the measurement system (Bond & Fox, 2007; Wright & Stone, 1979). Then, researchers need to decide if misfitting items need to be revised or omitted. However, this decision should be based on all available sources of evidence, including considerations that takes the context and where the evaluation tool is used into account (Schmeiser & Welch, 2006).

Another important aspect of validity pertains to evidence in relation to other variables such as group membership (e.g., world region, diagnosis). Not only can performances of groups be compared in terms of DIF (e.g., if item difficulties remain stable across world regions), but also in terms of a total test measure (e.g., if mean ADL motor or mean ADL process ability measures differ across world regions or diagnostic groups) (AERA et al, 1999; Camilli, 2006). If group differences arise, it is important to determine if they are grounded in genuine group differences in performance or from test or measurement bias (Camilli, 2006).

As I mentioned earlier, the AMPS software provides reports that allow for a norm-based interpretation of a person’s mean ADL motor and mean ADL process ability measures. Other evaluation tools—especially for use with children—also provide norms (e.g., the PEDI-CAT, [Haley et al., 2011]; the PEDI, [Haley et al., 1992]; or the M-ABC, [Henderson, et al., 2007]). Thus, when validity evidence in relation to world region is considered from the perspective of interpreting a child’s ADL task performance in reference to age-norms, it becomes important to evaluate if the mean ADL motor and mean ADL process ability measures of the AMPS remain stable. Differences in means among regional groups would raise concerns because they might have an impact on test fairness and the use of the AMPS with a specific group of persons (e.g., children from Middle Europe) (AERA et al., 1999). The current age norms of the AMPS were developed and calculated on the basis of mean ADL motor and ADL process ability measures for healthy, well persons across world regions, including children (Fisher & Merritt, 2012b). However, when evaluating norms, geographical region should always be taken into account (Crocker & Algina, 2008; Fawcett, 2007; Hammell, 2004). Despite the common agreement that the sample on which the norms are based should be large enough to provide stable values, no standards exist for defining good norms and often a sample of 30 participants per cell (e.g., participants from a specific region of a specific age) is used (Anastasi & Urbina, 1997; Bridges & Holler, 2007; Crocker & Algina, 2008; Fawcett, 2007). Only, Bridges and Holler (2007) provide a more clear recommendation based on an analysis of four pediatric psychological evaluation tools; that is, a sample size of 50 to 70 children per cell would be optimal to establish standardized z scores.
Introduction

Validity can also pertain to evidence in relation to diagnoses. That is, for example, the sensitivity of test scores for discriminating between children with and without disabilities. From a clinical perspective, such evidence is relevant for the use of the AMPS with children and the interpretation of the AMPS measures to identify those children who have ADL task performance problems, indicating a need for occupational therapy services. For evaluating evidence in relation to diagnoses, different approaches can be used. For example, linear statistical models and $t$ tests can be used to analyze if mean ADL motor and mean ADL process ability measures of children among diagnostic groups differ significantly or not.

Furthermore, it is strongly recommended to evaluate whether the differences among groups are of meaningful clinical importance (Cohen, 1988; Cumming, 2012; Ellis, 2010). Therefore, effect sizes can be evaluated to determine the magnitude of differences between groups and allow for interpretation of clinical meaningfulness. There are several approaches for evaluating effect sizes (Cumming, 2012). Probably the most common approach for evaluating effect sizes is to calculate Cohen's $d$ as an index of how much two groups differ when compared with their pooled $SD$s (Cohen, 1988). A $d$ value of 0.2 is considered to be a small effect size, meaning that the standardized mean difference between groups has no practical and clinical importance; an effect size of 0.5 is considered to be moderate, meaning that the difference is large enough to be visible to the naked eye; and an effect size of 0.8 is considered to be large, meaning the standardized mean difference between groups is of crucial practical and clinical importance. In some earlier studies that evaluated validity evidence of the AMPS measures, $SE$s were used as criteria to determine if the magnitude of differences between groups were of meaningful clinical importance (Fisher & Merritt, 2012b; Peny-Dahlstrand, Gosman-Hedström, & Krumlinde-Sundholm, 2012; Sperens, Munkholm, & Fisher, 2012).

Another aspect of the concept of validity is evidence related to the consequences of testing and test fairness for specific persons in specific groups. AERA et al. (1999) distinguish between evidence that is directly related to validity evidence from different sources and evidence that informs decisions, but falls outside the scope of validity. In the following, I will narrow the focus of evidence to consequences of testing that are relevant to my studies.

In general, evaluation tools are administered in the expectation that some beneficial information will be gained from the results (AERA et al., 1999). For occupational therapy, this information could be the basis for decisions regarding if the person is in need of intervention, intervention planning, evaluating the effectiveness of occupational therapy services, and/or the termination of intervention. Information about consequences of testing are particularly important when an evaluation tool is used in practice and the interpretation of the results leads to clinical decisions related to different groups of persons and performance differences are
Introduction

observed (AERA et al., 1999). The results of evaluation tools should be comparable across various groups (e.g., children from different world regions) and should enable fair comparisons and decisions based on results (e.g., if a child is in need for occupational therapy services or not). If the results differ from one group compared to another, then validity in relation to test fairness of an evaluation tool is questioned (AERA et al., 1999; Kane, 2006). In the case of occupational therapy, where decisions are central to the rationale for using evaluation tools, the direct examination of test fairness assumes high importance.

Validity evidence relation to test fairness can usually be evaluated in the context where the evaluation tool is used (Camilli, 2006). Evaluation of test fairness includes the evaluation of fairness a) as lack of bias (i.e., DTF); b) as equitable treatment in the testing process; c) as equality in outcomes of testing; and d) as opportunities to learn (AERA et al., 1999).

One approach for evaluating sources of validity evidence in relation to the consequences of testing and test fairness is to implement a feasibility study. In general, feasibility studies aim to assess methods and procedures of planned larger randomized controlled trials (RCTs) and their potential for successful implementation (Leon, Davis, & Kraemer, 2011; Thabane et al., 2010; Tickle-Degnen, 2013). This includes, among others, the evaluation of evaluation tools and data collection procedures used in order to refine procedures and avoid negative consequences in larger studies (Gardner, Gardner, MacLellan, & Osborne, 2003; Leon et al., 2011; Thabane et al., 2010). Feasibility studies are not implemented to test intervention hypotheses (Gardner et al., 2003; Leon et al., 2011; National Institute for Health Research Evaluation & Trials and Studies Coordinating Centre, 2011; Tickle-Degnen, 2013), or to estimate effect sizes (Leon et al., 2011). Recently, Thabane et al. (2010) provided detailed criteria for evaluating the success of these studies for the field of medicine, based on the primary feasibility objectives that can grouped under assessment of process, resources, management, and the scientific basis (van Teijlingen & Hundley, 2002; van Teijlingen, Rennie, Hundley, & Graham, 2001). Tickle-Degnen (2013) adapted and further developed these criteria for use in occupational therapy research.

Evidence Supporting Validity of the AMPS

The AMPS was developed based on the need for an objective evaluation tool focused on quality of performance of meaningful daily life tasks (Fisher & Merritt, 2012a). Many authors from all over the world have addressed validity evidence from different sources of the ADL motor and ADL process ability measures and scales of the AMPS (see Fisher & Merritt, 2012b for a summary), but it would go beyond the scope of this thesis to describe them in depth. In the following section I will narrow the focus of my summary to the validity evidence that is most relevant to my
studies, including evidence related to variables I chose not to control for (e.g., gender).

Evidence Related to Internal Structure

When I planned this PhD thesis, there was a lack of validity evidence related to the internal structure and stability of item difficulty calibration values of the AMPS across world regions, especially for use in Middle Europe. Earlier studies had evaluated item difficulty calibrations values of the AMPS across world regions and had found that overall the item difficulty calibration values remained stable across world regions (Fisher & Merritt, 2012b; Goldman & Fisher, 1997; Magalhães, Fisher, Bernspang, & Linacre, 1996). More specifically, Goldman and Fisher (1997) compared the item difficulty calibration values of North America, Nordic Countries, and the United Kingdom and found that none of the ADL motor or ADL process items displayed DIF. Magalhães et al. (1996) compared the item difficulty calibration values of a Swedish and a North American sample and found that only one of the ADL motor items (Endures) and none of the ADL process items displayed DIF. Similarly, in a recent study, Fisher and Merritt (2012b) found that only the ADL motor item Endures displayed DIF when they compared the item difficulty calibration values of North America, United Kingdom and Republic of Ireland, Nordic countries, all other European countries (including Eastern, Western, Southern, and Middle European countries), Australia and New Zealand, and Asian countries. None of these studies examined the stability of the item difficulty calibration values for Middle Europe as a separate world region.

Evidence Based on Relations to Other Variables

In terms of validity, other variables include variables external to the test (e.g., group membership in terms of age, world region, diagnosis, and/or gender) (AERA et al., 1999). When age and region are considered, there is limited validity evidence of the AMPS measures in relation to the use of the AMPS with children across world regions. Only one earlier study has evaluated for cross-regional differences in ADL ability (Peny-Dahlstrand et al., 2012). In that study, the mean ADL motor and ADL process abilities for children from North America and the Nordic Countries were compared to each other. The authors found a statistically significant region effect for ADL motor ability, but no significant region effect for ADL process ability. Moreover, when they did between region post hoc comparisons by age group, none of the mean differences were greater than 1 SE for the ADL motor and the ADL process ability measures of the AMPS. They also compared the Nordic means to the international age-normative means of the AMPS, and none of those differed by more than 1 SE from the international means. They concluded, therefore, that the AMPS remained free from cross-regional bias and that the international
age-normative means could be used in clinical practice in North America and Nordic countries (Peny-Dahlstrand et al., 2012).

Moreover, when I planned the studies of my PhD thesis, there existed some evidence of “diagnostic” group differences among children. For example, the results of earlier studies provided promising indication that the AMPS measures might be sensitive enough to identify differences between typically-developing children and children with and at risk for mild disabilities (Gol & Jarus, 2005; White & Mulligan, 2005; White et al., 2007). More specifically, Gol and Jarus (2005) compared 24 children without and 27 children with ADHD, 5 to 8 years, and found a significant difference in mean ADL process ability between groups, but no significant difference in mean ADL motor ability. When White and Mulligan (2005) compared ADL task performance in 21 children without a diagnosis and 12 children with ADHD (5 to 13 years), and White, Mulligan, Merrill, and Wright (2007) compared ADL task performance in 30 children without a diagnosis and 38 children with sensory processing deficits (5 to 13 years), they found significant differences in both mean ADL motor and ADL process ability between groups. In conclusion, these earlier studies revealed conflicting results and were based on small sample sizes. Therefore, there remained a need to cross-validate their results based on a larger sample in order to evaluate if the AMPS measures are sensitive enough to discriminate between typically-developing children and children with and at risk for mild disabilities.

When gender has been considered, research with the AMPS has provided evidence of lack of gender bias. More specifically, studies that investigated whether gender had an influence on validity of the AMPS measures found that there were significant differences on the ADL process scale that were practically and clinically not important because the magnitude of differences did not exceeded the mean SE of the AMPS measures (e.g., Fisher & Merritt, 2012b; Merritt & Fisher, 2003).

Evidence Related to Consequences of Testing and Test Fairness

When the existing validity evidence of the AMPS is considered in relation to consequences of testing and test fairness, the following conclusions can be drawn: the AMPS ADL tasks, the ADL motor and ADL process measures and scales have generally been shown to be free of gender bias (e.g., Merritt & Fisher, 2003) and free of DTF (e.g., Fisher & Merritt, 2012b). The findings of some studies do indicate a potential risk of cross-regional bias, especially when AMPS raters do not consider their own and the person’s cultural background when choosing ADL tasks for the AMPS observation and/or when rating certain ADL items on the AMPS (e.g., Uses) (Clawson, 1995; Goto, Fisher, & Mayberry, 1996). Furthermore, despite the differences in item difficulty calibration values among diagnostic groups, there is no evidence of systematic test bias (DTF) of the AMPS measures in relation to diagnoses (e.g., Fisher & Merritt,
Introduction

With the in-depth description of standardized administration procedures and the use of AMPS training courses (Fisher & Jones, 2012a, 2012b), the developers of the AMPS provide the basis for fairness as equitable treatment in the evaluation process (AERA et al., 1999; Camilli, 2006). However, there is evidence of difficulties in implementing the standardized administration procedures of the AMPS in clinical practice (e.g., Chard, 2000, 2004; Hitch, 2007) that might have an impact on consequences of testing and test fairness, for example, in terms of unequal treatment of person’s whose performance is evaluated (AERA et al., 1999; Camilli, 2006).

Considered together, there is evidence that supports the interpretations of the AMPS measures and scales with different diagnostic groups and in different world regions. However, additional questions related to the proposed use of the AMPS in typically-developing children and children with and at risk for mild disabilities and in persons from Middle Europe remained to be addressed.
RATIONALE

Occupational therapists promote peoples’ abilities to perform daily life tasks they need, want, or are expected to perform for their fullest possibly participation in society (Fisher, 2009). Therefore, occupational therapists need to evaluate people’s strengths and weaknesses with occupational performance in different areas of life (AOTA, 2002, 2008) so as to be able to set client-centered and occupation-focused goals, to design and implement occupation-based and/or occupation-focused interventions, and to evaluate the effectiveness of their services (Fisher, 2013).

Contemporary occupational therapy practice demands the use of standardized evaluation tools (Law et al., 2005) that inform decisions (AERA et al., 1999) about people and occupational therapy intervention programs. Thus, an improper use of evaluation tools can have considerable consequences for people whose occupational performance is evaluated and for occupational therapy in general. The use of evaluation tools in a sound and ethical way is crucial for occupational therapist working in clinical practice and/or research. Both, practitioners and researchers (i.e., test developers) share the responsibility for providing validity evidence from different sources that supports the use of evaluation tools in a specific context or a specific group of people (AERA et al., 1999).

In the context of Middle Europe, there has been a gap of valid and reliable evaluation tools (Jobst, 2008; Siebert et al., 2007), including occupation-based and/or occupation-focused tools that can be used with the specific group of children with and at risk for mild disabilities (Projektgruppe ergotherapeutische Befundinstrumente in der Pädiatrie, 2004; Röse & Seitz, 2008). The AMPS was one standardized evaluation of occupational performance that has been used and validated, internationally (Fisher & Jones, 2012a, 2012b). An increasing number of occupational therapists in Middle Europe have started to use the AMPS in their daily practice for evaluating people’s quality of ADL task performance. Thus, the AMPS needed to be evaluated in this “new context” (AERA et al., 1999). Furthermore, there has been limited validity evidence of the AMPS when used with children with and at risk for mild disabilities. Earlier studies have provided promising evidence that the AMPS might be valid for this group of people, but the results of these studies had been based on small sample sizes and revealed conflicting results (Gol & Jarus, 2005; White & Mulligan, 2005; White et al., 2007).

There remained a need, therefore, to generate additional evidence to support the use of the AMPS in occupation-based and occupation-focused evaluation and intervention, especially in the context of Middle Europe and with children with and without mild disabilities.
Aims

AIMS OF THE THESIS

The purpose of this thesis was to contribute evidence to support the valid use of the AMPS in occupation-based and occupation-focused evaluation and intervention with typically-developing children and children with and at risk for mild disabilities, including children living in Middle Europe.

Specific Aims

To evaluate for evidence to support the validity of the AMPS related to internal structure—stability of item difficulty calibration values for a Middle European sample compared to samples from other world regions (DIF) (Study I).

To evaluate for evidence to support the validity of the AMPS measures in relation to other variables—stability of the mean AMPS measures of typically-developing children from Middle Europe compared to mean AMPS measures of typically-developing children from other world regions (Study II).

To evaluate for evidence to support the validity of the AMPS measures in relation to other variables—sensitivity of the AMPS measures for discriminating between typically-developing children and children with and at risk for mild disabilities (Study III).

To evaluate for evidence to support the validity of the AMPS measures in relation to consequences of testing and test fairness—determination of whether or not any identified DIF results in DTF (Study I).

To evaluate for evidence to support the validity of the AMPS measures in relation to the use of the AMPS as a standardized, occupation-based and occupation-focused evaluation tool in the context of a feasibility study (Study IV).
METHODS

Participants

The participants in this thesis consisted of groups of persons of different ages, from different world regions, with and without diagnoses. In Study I, we aimed to evaluate cross-regional validity of the AMPS, we included persons from Middle Europe and other world regions, all ages, and with and without diagnoses. For Studies II to IV, we only included typically-developing children and/or children with and at risk for mild disabilities, from Middle Europe, and other world regions. I provide an overview of participant’s demographics in Table 2 and give specific information related to the four studies below.

Study I

Eligible participants for Study I, were all available persons from the international AMPS database, Ft. Collins, Colorado, USA as of June 2010 who had been scored by raters in a valid manner (i.e., free of rater scoring error). They were 145,489 persons, 3 years of age and above, thereof 1,346 were persons from Middle Europe. They included well persons and those with a broad variety diagnoses (e.g., orthopedic/musculoskeletal, neurologic). They reflected the variety of persons evaluated by occupational therapists in Middle Europe and internationally. The 144,143 persons who were not from Middle Europe were from North America, United Kingdom/Ireland, the Nordic Countries, other Europe, Australia/New Zealand, and Asia (see Table 2). The data for the Middle European sample had been submitted to the AMPS database by 117 occupational therapists who have attended AMPS courses in Austria, Germany, Slovenia, and Switzerland, and who calibrated as reliable and valid AMPS raters. Sample size selection for this study was based on the premise that a minimum of 200 persons are required in each of the regions (i.e., Middle Europe and each of the other world regions), but that it is desirable to use the largest possible sample sizes when performing DIF analyses (Zumbo, 1999). For the purpose of this study, we included data for persons from Austria, Germany, Slovenia, and Switzerland in the Middle European sample as AMPS data only were available for those Middle European countries. We are aware, however, that Middle Europe can be considered to extend beyond the borders of those four countries.
### Table 2

**Demographic Data: Number of Participants, Age, Gender, World Region, and Diagnostic Characteristics of Participants**

<table>
<thead>
<tr>
<th>Study</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of participants</td>
<td>145,489</td>
<td>11,189</td>
<td>10,996</td>
<td>17</td>
</tr>
<tr>
<td>Age (years)</td>
<td>3-103</td>
<td>2-15</td>
<td>4-15</td>
<td>5-9</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>54.1 (24.4)</td>
<td>8.1 (3.4)</td>
<td>8.7 (3.2)</td>
<td>7.1 (1.4)</td>
</tr>
<tr>
<td>Gender (percentage)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>63,860 (43.9)</td>
<td>5,012 (44.7)</td>
<td>5,328 (48.5)</td>
<td>11 (64.7)</td>
</tr>
<tr>
<td>Female</td>
<td>81,576 (56.1)</td>
<td>6,165 (55.1)</td>
<td>5,668 (51.5)</td>
<td>6 (35.3)</td>
</tr>
<tr>
<td>Unknown</td>
<td>53 (0.0)</td>
<td>12 (0.1)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>World Region (percentage)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North America</td>
<td>23,441 (16.1)</td>
<td>2,607 (23.3)</td>
<td>2,422 (22.0)</td>
<td>-</td>
</tr>
<tr>
<td>United Kingdom / Republic of Ireland</td>
<td>29,670 (20.4)</td>
<td>2,420 (21.6)</td>
<td>2,142 (19.5)</td>
<td>-</td>
</tr>
<tr>
<td>Nordic countries</td>
<td>47,445 (32.6)</td>
<td>2,805 (25.1)</td>
<td>2,931 (26.7)</td>
<td>-</td>
</tr>
<tr>
<td>Other Europe</td>
<td>15,057 (10.4)</td>
<td>577 (5.2)</td>
<td>683 (6.2)</td>
<td>-</td>
</tr>
<tr>
<td>Australia/New Zealand</td>
<td>13,329 (9.2)</td>
<td>933 (8.3)</td>
<td>899 (8.2)</td>
<td>-</td>
</tr>
<tr>
<td>Asia</td>
<td>15,201 (10.4)</td>
<td>1,569 (14.0)</td>
<td>1,509 (13.7)</td>
<td>-</td>
</tr>
<tr>
<td>Middle Europe</td>
<td>1,346 (0.9)</td>
<td>278 (2.5)</td>
<td>264 (2.4)</td>
<td>17 (100)</td>
</tr>
<tr>
<td>South and Central America and Caribbean</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>11 (0.1)</td>
</tr>
<tr>
<td>Middle East</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>84 (0.8)</td>
</tr>
<tr>
<td>Africa</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>49 (0.4)</td>
</tr>
<tr>
<td>Diagnostic group (percentage)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Typically-developing, well</td>
<td>12,618 (8.7)</td>
<td>11,189 (100)</td>
<td>9,615 (87.4)</td>
<td>-</td>
</tr>
<tr>
<td>At risk for or with mild disabilities</td>
<td>1,304 (0.9)</td>
<td>-</td>
<td>-</td>
<td>1,383 (12.6)</td>
</tr>
<tr>
<td>Other</td>
<td>131,567 (90.4)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*Note:* Persons with other diagnoses in Study I were older adults 60 years of age and older, who were frail or at risk for functional decline, but without known medical problems; or persons with neurological developmental disorders, right- or left-sided cerebral vascular accident, other types of neurological disorders, intellectual disabilities, musculoskeletal disorders, medical conditions (e.g., cardiovascular, respiratory, burns, AIDS/HIV) or sensory disorders (e.g., visual, auditory, vestibular), other psychiatric disorders or disorders on the autism spectrum, two or more diagnoses from different categories or whose diagnoses were unknown, memory disorders not associated with dementia, or dementia, schizophrenia, other type of thought disorder.
Methods

Study II

For Study II, we selected participants from the international AMPS database in February 2013 if they met following inclusion criteria: a) were between 2 and 15 years of age from Middle Europe, North America, United Kingdom/Ireland, the Nordic countries, Australia/New Zealand or Asia; and b) were children without a known disability or diagnosis. The age of 2 was chosen as the lower boundary because the AMPS is used to evaluate persons 2 years and above. The age of 15 was selected as the upper boundary because ADL motor and ADL process abilities seem to increase with age until the age of 15 and then plateau between 16 and 50 (Hayase et al., 2004). Excluded were participants who had been scored by raters in an invalid manner.

Our initial plan for sample size was based on the premise that a minimum of 30 persons is required in each age group for each region to obtain stable estimates of group means (Eckes, 2011; Linacre, 1994, 2004). As we had less than 30 children in some groups, we selected a statistical model that could handle smaller sample sizes. The model is discussed in more detail below. We recognized that a larger sample size would have been better to estimate means for a Middle European sample of typically-developing children (Bridges & Holler, 2007; Eckes, 2011; Linacre, 1994, 2004). We compared the international means (Fisher & Jones, 2012a) with the means for the Middle European sample, knowing that our sample sizes were too small. A total of 11,189 children between 2 and 15 years ($M=8.1$, $SD=3.4$) met the criteria. Demographic data for the participants are presented in Table 2.

Study III

For Study III, we selected participants from the international AMPS database in May 2011 if they met the following inclusion criteria: a) were between 4 and 15 years of age; b) were children without a diagnosis, children at risk (i.e., not diagnosed, but having symptoms suggestive of having a mild disability), or children diagnosed with mild disabilities (i.e., ADHD, DCD, LD, and/or SI); and c) had no other neurological disorders (e.g., traumatic brain injury, cerebral palsy), pervasive developmental disorder, mental health problem (e.g., childhood depression), or intellectual disability. The age of 4 was chosen as the lower boundary because while the AMPS database includes data for persons 2 years and above, there were insufficient data for children with mild disabilities aged 2 and 3. The age of 15 was selected as the upper boundary because both ADL motor and ADL process ability increase with age until age 15 years and then plateau between 16 and 50 years (Hayase et al., 2004). Excluded were children who had been scored by raters in an invalid manner (i.e., were not free of rater scoring error which can occur when raters are unexpectedly lenient or strict). For more details see Fisher and
Methods

An analysis, with an anticipated effect size of 0.80 or higher, an $\alpha$ of 0.05, and power of 0.80 revealed a needed sample size of 21 in each group for the probability of finding significant group differences when differences really exists. We did not match the group of children with mild disabilities to the group of children without a diagnosis because of the preferred premise of using the largest possible sample (Zumbo, 1999). A total of 10,998 children between 4 and 15 years ($M=8.7, SD=3.2$) met the criteria. There were 5,328 (49%) boys and 5,668 (51%) girls, and 9,615 (87%) were children without a diagnosis and 1,383 (13%) were children at risk for or with mild disabilities. The participants were from a variety of world regions (see Table 2).

Study IV

Eligible participants for this study were all children a) 5 to 9 years of age, b) who had been diagnosed by pediatricians with ADHD and/or DCD, according to of DSM-IV criteria (APA, 2000) within the last 6 months, c) who had problems with ADL task performance based on parental report; d) were eligible for occupational therapy services, and e) whose parents agreed to participate in the study. The age of 5 was chosen as a minimum age because ADHD and DCD are mostly diagnosed beyond this age (APA, 2013; Lahey et al., 1994). The age of 9 years was chosen as the upper limit because in Switzerland, children have to be diagnosed with ADHD by this age (Der Schweizerische Bundesrat, 2004), otherwise, the main insurance company (Swiss Insurance for Disabled and Elderly Citizens) will not fund occupational therapy services. Therefore, most pediatricians strive to make the diagnosis before the age of 9. Exclusion criteria were a diagnosis of neurological disorder (e.g., traumatic brain injury or cerebral palsy), autism spectrum disorder, other mental disorders (e.g., childhood depression), intellectual disabilities, or social problems (e.g., parents were not able to bring the child to the weekly occupational therapy sessions). The eligible children and their parents were recruited by 10 pediatricians and 15 occupational therapists. The pediatricians and occupational therapists were informed verbally and in writing about the study. Then, they spoke to children and their parents about the study and provided written invitations describing the research procedures.

Between February 2011 and February 2012, 29 children were found to be eligible participants, and 17 of them were included in the study. The flowchart of participants is presented in Figure 3.
Methods

Figure 3. Flow chart of participants in Study IV
Methods

Evaluation Tools and Procedures

We used the AMPS (Fisher & Jones, 2012a, 2012b) to evaluate observable quality of ADL task performance of all persons in Study I and all children in Studies II, III and IV. In Study IV, we additionally used the COPM (Law et al., 2009) to evaluate perceived problems in quality of occupational performance and satisfaction. An overview of evaluation tools and methods used for data collection is presented in Table 3.

<table>
<thead>
<tr>
<th>Study</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation tool</td>
<td>AMPS</td>
<td>AMPS</td>
<td>AMPS</td>
<td>AMPS COPM</td>
</tr>
<tr>
<td>Data collection method</td>
<td>Observation</td>
<td>Observation</td>
<td>Observation</td>
<td>Minutes of meetings, e-mails, documentation of AMPS and COPM evaluations, intervention protocols, field notes</td>
</tr>
<tr>
<td>Design</td>
<td>Retrospective</td>
<td>Retrospective</td>
<td>Retrospective</td>
<td>Prospective</td>
</tr>
</tbody>
</table>

Use of the AMPS in Studies I to IV

Observed quality of ADL task performance of all participants was evaluated using the AMPS according to the standardized administration procedures described in the AMPS manuals (Fisher & Jones, 2012a, 2012b). These procedures were implemented by occupational therapists who were trained and calibrated AMPS raters. Each occupational therapist started with preparing for an AMPS interview for a specific person, including considerations of what was the range of appropriate ADL tasks options to offer the person. Then, the occupational therapist interviewed the person and/or parents or teachers about the daily life tasks the person wanted, needed, and/or was expected to perform. The person and the occupational therapist then decided together which two tasks the person was going to perform during the AMPS observation based on information gathered during the interview. Afterwards, the person and occupational therapist became familiar with the environment where the person was evaluated. The setting could be a natural
environment or a clinical setting that was as naturalistic as possible. Then, the occupational therapist summarized the shared decisions and initiated the ADL task observation where the person performed at least two AMPS tasks. During each ADL task performance, the occupational therapist observed the person and took notes. Only after the person had finished did the occupational therapist score the person’s quality of ADL task performance based on his or her notes. He or she scored 36 ADL motor and ADL process items in terms of any observed increase in physical effort or clumsiness, decrease in efficiency, decrease in safety, or frequency of assistance provided in relation to each ADL skill (smallest observable unit of the overall ADL task performance). Every AMPS item was rated using a four point ordinal scale and the raw scores were entered into the AMPS software (AMPS Project International, 2010) that can be used to generate AMPS Graphic and Summary Reports of the person’s actual quality of ADL task performance. The occupational therapist then interpreted the results of the AMPS observation for further decision making (e.g., need of intervention, goals for intervention). Further descriptions and evidence of validity of the AMPS are presented in the Introduction.

Use of the COPM in Study IV

In Study IV, the Canadian Occupational Performance Measure (COPM) (Law et al., 2009) was used as a second evaluation tool. The COPM is a client-centered evaluation tool designed for use by occupational therapists to detect change in a person’s self-perception of occupational performance and the satisfaction with this performance over time (Law et al., 2009). The COPM evaluation started with a semi-structured interview with parents who were asked to identify issues in the areas of self-care, productivity (e.g., school) and leisure. Once the parents identified problems in daily activities, they rated the importance of each activity on an ordinal scale from 1 to 10. From this list, the parents choose between two and five problems they wished to focus on. For each of the problems, the parents then rated performance and satisfaction with performance, again using an ordinal scale from 1 to 10. Higher ratings indicated better perceived quality of performance and satisfaction with the activity. The performance and satisfaction scores of the selected daily activities were summed and averaged over the number of problems. Differences between baseline and subsequent scores indicated the outcome. A change score of 2 or more is considered to be clinically meaningful (Law et al., 2009).

The COPM was tested for the use with children and was identified as being able to be used to detect unique child-specific occupational performance problems (Cup, Scholte op Reimer, Thijssen, & van Kuyk-Minis, 2003; Cusick, Lannin, & Lowe, 2007; Cusick, McIntyre, Novak, Lannin, & Lowe, 2006; Kjeken, Slatkowsky-Christensen, Kvien, & Uhlig, 2004; Pan, Chung, & Hsin-Hwei, 2003; Verkerk, Wolf, Louwers,
Methods

Meester-Delver, & Nollet, 2006). It is recommended that under the age of 8, parents rather than children should complete the COPM because children have problems identifying activities where they experience difficulties (Law, 2012). It has been shown that the COPM has been successfully used with parents as proxy-reporters (Cusick et al., 2007). In Switzerland the COPM is widely recognized and in use. Moreover, the Swiss Association of Occupational Therapists has advocated for the use of the COPM as an evaluation tool in order establish quality assurance in client-centered occupational therapy practice (EVS, SRK, & Santesuisse, 2007). Hence, the COPM was used as the second evaluation tool in Study IV.

Procedures in Study IV

Settings and Location

The study took place in Switzerland, at a center for child neurology, development, and rehabilitation and six occupational therapy private practices in Swiss communities. For intervention, eligible occupational therapy practitioners in a specific region were selected based on following inclusion criteria: a) they had a minimum of 3 years of clinical experience, b) knowledge and skills in working with children with ADHD and/or DCD and their families, c) had knowledge and skills in using standardized evaluation tools, d) were willing to participate in a 3-day training course (where they were introduced to the Occupational Therapy Intervention Process Model (OTIPM) (Fisher, 2009), and e) were willing to participate in regular meetings with the research team. In total, eight occupational therapists implemented the occupational therapy intervention following the study protocol.

The occupational therapist who served as a blind evaluator was selected based on a) having a minimum of 3 years of clinical experience and b) training as a valid and reliable AMPS rater. The evaluator was not part of the intervention team and was not informed as to which group the participants were allocated. She rated the AMPS three times per child.

Intervention

After baseline, participants were randomly assigned to receive either occupational therapy intervention or to a waiting time control condition. The occupational therapy intervention followed a study protocol based on the principles of the OTIPM (Fisher, 2009) which is a client-centered, occupation-based, and occupation-focused approach. The OTIPM has recently been used in other intervention studies (e.g., Simmons & Griswold, 2010). The intervention process, guided by the professional reasoning proposed in the OTIPM, started with informing the child and his or her parents about occupational therapy as relevant to this
Methods

intervention study. Then, the occupational therapist interviewed the child, his or her parents and/or important others (e.g., teachers). Interview questions addressed issues such as child’s strengths and problems in daily activities; his or her roles; habits and interests; and potential facilitators and barriers in the physical, attitudinal, and social environment. Then, the occupational therapist evaluated the quality of the child’s ADL task performance in using the AMPS. Afterwards, a task analysis was implemented in order to define and interpret the factors that contributed to the child’s diminished quality of occupational performance (e.g., environmental, personal factors that influenced the child’s problems). The occupational therapist formulated measurable goals in collaboration with the child, his or her parents and/or important others, based on the interviews, the child’s and his or her parents priorities, and the results of the baseline evaluation. At least one goal was formulated in the area of ADL (i.e., self-care and domestic life). The following intervention focused on enhancing child’s performance in meaningful and purposeful daily activities and was implemented in weekly sessions, 60 minutes each, over 15 weeks. Participants in the control group did not receive any intervention during the 15 week waiting time period, but were provided with the same occupational therapy intervention after the control phase. We decided not to compare this occupational therapy intervention to a placebo, because, to our knowledge, no placebo for occupational therapy has achieved successful blinding of children and their parents (e.g., parents talk to each other in the waiting room) and because of political issues (i.e., this is one of the first occupational therapy intervention studies implemented in Switzerland).

Evaluation tools

The AMPS (Fisher & Jones, 2012a, 2012b) was used as the objective (“outsider” view of observed performance) and primary evaluation tool. The COPM (Law et al., 2009) was used as a secondary evaluation tool (insider view of perceived performance) of children’s parents. AMPS and COPM were administered three times at baseline, post-test 1/pre-test 2, and post-test 2/post-test 1 (see Figure 3). Ideally, a post-test after 6 months would have been administered, but we decided that it was not ethically sound to withhold treatment to the control group for 6 months. The AMPS was video-rated by the blinded rater, the COPM was completed by unblinded raters. While the COPM was administered, and its use supported selection of which AMPS tasks to observe, its use was not otherwise related to the focus of this PhD thesis, and therefore, I do not discuss it further in the frame of this thesis.

Randomization

Independent persons, who were not otherwise involved in the study and had no detailed information about it, allocated participants randomly to the intervention or control group. A random block design was used
Methods

because enrollment occurred over an extended period of time (Polit & Beck, 2008).

Blinding

In an intervention study like this, it is impossible to blind participants, their parents, pediatricians, and occupational therapy practitioners for group allocation. However, blinding was strictly maintained for researchers and the AMPS rater. During the data collection period, we maintained separation between the blinded AMPS rater and the occupational therapy practitioners who delivered the interventions. To ensure blinding to group assignment, the AMPS observations were videotaped by intervening occupational therapists and rated by the blinded AMPS rater.

Qualitative Data Collection Procedure

We collected qualitative data to evaluate feasibility during the whole study. The data collection process started at the first meetings where occupational therapy practitioners, together with the research team, planned the study. The data collection process ended with member-check, where we discussed the emerging data and interpretations with the practitioners. The raw data consisted of a) minutes of meetings between researchers, pediatricians, doctors of the insurance companies, and occupational therapy practitioners (who did intervention or the blinded rating); b) minutes of telephone meetings and e-mails; c) documentation of AMPS and COPM evaluations; d) interventions protocols; and e) my field notes. The minutes of meetings were written by myself, checked by my Swiss supervisor, and distributed to all persons involved in the meetings. The field notes were written after every meeting and included reflections by myself, distributed to and discussed with my supervisors in Sweden, Switzerland, and the USA.
Data Analysis

We used different methods to analyze the data from the four studies included in this PhD thesis. An overview of methods of data analyses is provided in Table 4.

Table 4
Overview over Data Analysis Methods in Studies I to IV

<table>
<thead>
<tr>
<th>Study</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive statistical analyses</td>
<td>$M, SD$, range, percentage</td>
<td>$M, SD$, range, percentage</td>
<td>$M, SD, Mdn$, range, percentage</td>
<td>$M, SD$, range, percentage</td>
</tr>
<tr>
<td>Rasch analyses methods</td>
<td>MFR, DIF analyses</td>
<td>MFR analyses</td>
<td>MFR analyses</td>
<td>MFR analyses</td>
</tr>
<tr>
<td>Inferential statistical analyses</td>
<td>DTF analyses</td>
<td>Two-way ANOVAs, contrast estimation</td>
<td>Forced regression, post-hoc $t$ tests</td>
<td>-</td>
</tr>
<tr>
<td>Effect sizes</td>
<td>Difference of $\pm 0.55$ logit</td>
<td>Difference of $\pm 1$ $SE$</td>
<td>Cohen’s $d$</td>
<td>-</td>
</tr>
<tr>
<td>Qualitative analysis</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Deductive content analysis</td>
</tr>
</tbody>
</table>

Descriptive Statistical Analyses

We implemented descriptive statistical analyses across all four studies (see Table 4). We used the statistical software IBM SPSS Statistics, Version 20.0 (SPSS, 1989-2011).

Rasch Analysis Methods

In Studies I to III, we used the FACETS (Linacre, 1987-2012) computer program that allowed us to handle large datasets and to implement a variety of MFR analyses. More specifically, in Study I, we unanchored the item difficulty facet in order to estimate item difficulty calibration values for each world region in order to implement DIF analyses for the ADL motor items and for the ADL process items (i.e., calculating the logit differences between world regions). In total, we compared Middle Europe to six other world regions as well as to the total sample, resulting in seven comparison pairs.
Methods

In Studies II and III, we used FACETS to generate the ADL motor and ADL process ability measures for each participant and to convert ordinal ADL motor and ADL process item raw scores into two linear measures. In Studies II and III we performed two MFR analyses, one to generate the ADL motor ability measures and one to generate the ADL process ability measures. In each analysis, each rater’s severity, each ADL task challenge, and each item difficulty was anchored at pre-established values based on those in the current AMPS software (AMPS Project International, 2010).

In Study IV, the AMPS software was used to generate linear ADL motor and ADL process ability measures of the participants (AMPS Project International, 2010).

Inferential Statistical Analyses

We employed inferential statistical analyses in Studies I to III using SPSS (SPSS, 1989-2011). In Studies II and III, we used linear statistical models for the main data analyses and we set the level of significance at $\alpha=.05$. Study specific analyses are discussed below.

Study I

In Study I, after the DIF analyses, we evaluated DTF by plotting the ADL measures for all persons when estimated based on the item difficulty calibrations for Middle Europe against the ADL measures for all persons when based on the item difficulty calibrations for each of the other world regions or the total combined sample. We then evaluated if the paired ADL measures fell within a 95% confidence interval, indicating no evidence of DTF (Bond & Fox, 2007; Wright & Stone, 1979). More specifically, two different sets of item difficulty calibration values are expected to yield invariant person measures such that the plotted measures fall within 95% control lines that are based on the SEs for each item pair (Bond & Fox, 2007; Wright & Stone, 1979).

Study II

In Study II, we evaluated cross-regional differences of mean ADL motor and mean ADL process ability measures of the AMPS among children from different world regions. Because our sample size for the Middle European age groups was lower than anticipated, we first evaluated which statistical model would best fit our data. An optimal model should have good fit (small RSS—residual sum of squares) and a simple structure (small number of parameters or variables). Therefore, we tested for the most parsimonious model; with age as a continuous variable, with cubic, quadratic, and linear terms; and with and without interaction effects. We determined which model was most parsimonious using the Akaike information criterion (AIC). The AIC is a value that penalizes complicated models and quantifies the optimality of a model, which means that the...
model has good fit with the data, but is not too complicated (Norman & Streiner, 2008). We then implemented two two-way ANOVAs to estimate age and region effects, one for mean ADL motor and one for ADL process ability measures, respectively. The following linear two-way model was fitted to the data:

\[ Y_{ijk} = \mu + \alpha_i + \tau_j + \gamma_{ij} + \epsilon_{ijk} \]

with \( k=1, \ldots, n_{ij} \), \( i=2, \ldots, 15 \), \( j=1, \ldots, 7 \), \( \mu \) as intercept, \( \alpha_i \) as the \( i \)-th age effect, \( \tau_j \) as the \( j \)-th region effect, \( \gamma_{ij} \) as \( ij \)-th interaction and \( n_{ij} \) as the age-region cell size. We implemented residual analyses to check model assumptions, that is, a) independence, b) zero-expectation, and c) variance homogeneity of the \( \epsilon_{ijk} \)’s.

Once we had performed the ANOVAs, we explored our results in more detail by estimating contrasts between Middle Europe and each of the six other world regions for each age based on the linear model. In total, we estimated 84 contrasts (14 ages x 6 world regions) for the ADL motor ability measures and 84 contrasts for the ADL process ability measures. We implemented simultaneous confidence interval procedures to control for the family-wise error rate, using multivariate normal and \( t \) distributions and Sidak corrections to avoid the risk of false positive results.

**Study III**

In Study III, we implemented two forced regression analyses to evaluate for significant differences in mean ADL motor ability measures and mean ADL process ability measures between groups. The regression model can be clarified with the following equation:

\[ Y_{i} = \beta_0 + \beta_{group_i} + \beta_{age_i} + \beta_{group_i*age_i} + \epsilon_i \]

where \( Y \) is the mean ADL motor or mean ADL process ability measure. Age, group, and group*age were the independent variables; \( \beta_0 \) the intercept; \( \beta_{i-3} \) the effects for group, age, and age*group interaction; and \( \epsilon \) the normally distributed error.

Finally, we tested for differences in ADL motor and ADL process ability measures between the groups by age category using post hoc one-tailed \( t \) tests. Levene’s tests were used to evaluate for equality of variances between groups. When the results of Levene’s tests failed to support the assumption of equality of variances between groups, the \( t \) values for samples where equality of variances is not assumed were used. Bonferroni multiple comparison corrections were used to avoid the risk of false positive results. This is known to be a conservative method for ensuring a familywise \( \alpha \) level of 0.05. With 12 comparisons each for ADL motor and ADL process, the critical value for \( \alpha \) was set at .004.
Methods

Effect Sizes

We calculated effect sizes in Studies I to III in order to evaluate if differences between groups (i.e., world region or diagnoses) were of clinical importance. We used different approaches to calculate effect sizes based on the research questions and inferential statistical models that we used for our main data analyses.

Study I

In Study I we had a very large sample, resulting in the risk of too much power and over identifying significant differences based on p values alone (Cumming, 2012; Wilson, 2005). Therefore, we used effect sizes to evaluate for significant DIF (Tristán, 2006; Wilson, 2005). More specifically, we set our criteria for the presence of significant DIF based on a logit difference of at least ±0.55 logit between Middle Europe and each of the other six world regions and the total combined sample. This criterion was based on Tristán (2006) who found that when standard errors (SEs) are normalized, the minimum possible SE is 0.20 logit. With SE values of 0.20 logit, a difference in item difficulty calibration values of ±0.55 logit is required for statistical significance. The rationale for the use of Tristán’s criterion is discussed in more detail by Munkholm, Berg, Lögren, and Fisher (2010), who implemented a similar cross-regional validation study of the School Version of the AMPS (Fisher et al., 2007).

Study II

In Study II, where we compared the international age-normative means of the ADL motor and ADL process ability measures of the AMPS to the Middle European means, we defined a meaningful difference as one equal to or greater than the mean SE for the ADL motor (0.25 logits) or the ADL process ability measures (0.20 logits) of the AMPS. This method has been implemented in prior studies (Fisher & Merritt, 2012b; Sperens et al., 2012).

Study III

In Study III, where we compared mean ADL motor and ADL process ability measures of typically-developing children and children with and at risk for mild disabilities across ages, we calculated Cohen’s d as an index of how much two groups differed when compared with their pooled SDs (Cohen, 1988). We calculated d as \[ [(\bar{x}_1 - \bar{x}_2)/\sqrt{(s_1^2 + s_2^2)/2}] \]. We used Cohen’s d because the SDs of the groups did not differ and the homogeneity of variance assumption was not violated (Ellis, 2010).
Descriptive Analysis of Feasibility

The primary purpose of Study IV was to evaluate feasibility of an occupational therapy intervention study in order to build up a basis for a larger randomized controlled trial (RCT). The evaluation of feasibility included the evaluation of the implementation of the AMPS (Fisher & Jones, 2012a, 2012b) as a new evaluation tool and the implementation of a new occupation-based and occupation-focused intervention based on the OTIPM (Fisher, 2009) in Swiss practice. In the following section of this thesis, the methods of the data analysis will be described as a whole and the results of Study IV will be described from a validity perspective, with the use of the AMPS as an evaluation tool in context of a feasibility study.

We analyzed data based on the principles of content analysis (Mayring, 2010). We defined the feasibility objectives proposed by Thabane et al. (2010) and Tickle-Degnen (2013) as main categories for our coding frame. More specifically, assessing feasibility of the process, resources, management, and scientific basis of the planned larger RCT was the focus of our analysis. Furthermore, we identified subcategories (e.g., recruitment) that specified aspects that constitute the four main categories, both deductively in a theory driven way and inductively in a data driven way. More precisely, first we used the specific feasibility objectives (Thabane et al., 2010; Tickle-Degnen, 2013) to identify subcategories and coded our data accordingly. During coding, we revised and expanded subcategories by going back and forth between theory and data. An overview of the main categories and subcategories are presented in Table 5.

I coded the data based on the coding frame that all authors had agreed on. I discussed the allocation of the codes, the revision, and expansion of the subcategories with my Swiss supervisor. All other supervisors verified the data analysis in various discussions. Furthermore, we used peer debriefing with researchers at the Umeå University and at Zurich University of Applied Sciences (ZHAW) as well as member checking with occupational therapy practitioners for establishing credibility of our interpretations (Polit & Beck, 2008).
Methods

Ethical Considerations

Throughout the research process, we paid close attention to dynamic ethical issues. Our intention was to design and implement ethically sound research and take responsibility for our decisions. We considered general ethical issues when designing, implementing, and reporting our research (e.g., following publication rules) (Vetenskapsrådet, 2011) and specific ethical issues as we implemented our research across borders, with children as research participants, with data from an international database, and in a clinical trial. In the following section, I address these specific ethical considerations and related decisions of the research project.

Table 5
Overview over Main Categories and Subcategories of the Deductive Content Analysis in Study IV

<table>
<thead>
<tr>
<th>Main category</th>
<th>Subcategories</th>
</tr>
</thead>
</table>
| Process assessment | • Recruitment rate and randomization  
                       | • Retention rate  
                       | • Inclusion and exclusion criteria  
                       | • Data collection                                                             |
| Resources assessment | • Process time, for training and communication with stakeholders, supervision during the process, implementing new methods in a specific region  
                               | • Access to equipment and material  
                               | • Commitment to research                                                     |
| Management assessment | • Competences within research team  
                               | • Geographical distribution  
                               | • Research in a new setting and context  
                               | • Documentation of results  
                               | • Documentation of intervention                                               |
| Scientific basis assessment | • Study protocol  
                               | • Criteria for statistical significance and clinical meaningfulness  
                               | • Research design                                                             |

Note. The main categories are based on objectives of feasibility studies recommended by Thabane et al. (2010) and (Tickle-Degnen 2013). Encountered strengths and problems are the subcategories of our findings.
Methods

Given that I was a doctoral student at Umeå University in Sweden, but mainly conducted the research in Switzerland, we had to follow both Swedish and Swiss ethical regulations. For Studies I to III, where we had accessed pre-existing, anonymous data from the AMPS database, we had obtained ethical approval from the Regional Ethical Review Board, Faculty of Medicine, Umeå University, Sweden (Dnr03-509). Furthermore, the Ethics Committee of Canton Zurich confirmed that the secondary analysis of anonymous medical data did not need to be submitted to the Ethics Committee in Switzerland. For Study IV, which was implemented in the Swiss Cantons Appenzell, St. Gallen, and Thurgau, we received approval from the ethics board of the Canton St. Gallen, Switzerland, as it is responsible for these Eastern Swiss Cantons (EKSG 10/008/1B).

Research done with children as research participants needed to be considered very critically because children are a group of vulnerable persons (Brody, 1998). Research with children as research participants should generally only be developed when no other group of adults could be used to answer the research questions (Levine, 1986; Smith, 1999). Since the foci of Studies II to IV was on the validity of the AMPS and the feasibility of a planned RCT with children, only children could serve as research participants. Therefore, we built up our ethical decisions based on Brody (1998), who described issues that need to be considered when doing research involving vulnerable persons. In Study IV, we obtained informed consent from legal guardians and from the children as the research participants (World Medical Association, 1996). In the parents informed consent sheet, the study was described briefly and it was highlighted that participation was entirely voluntary and they could refuse to participate or withdraw from the study at any stage, with no further consequences. The children’s consent sheet was simplified and adapted to the age of the children. Furthermore, parents and their children were informed about the intended research project and all benefits and risks (principle of autonomy). Third, we maintained confidentiality and secrecy. Furthermore, we designed and constantly evaluated if the level of risk in the research (e.g., waiting time of the control group) was acceptable (principle of non-maleficence). Last, but not least, we concluded that our research was in the interest of the children with mild disabilities (principle of fairness or justice), that neither the observation of ADL behavior, nor the occupational therapy intervention was harmful for them, and that the children could directly benefit from the research (principle of beneficence). To summarize, informed consent was obtained from legal guardians and the children, they could decide to withdrawal from the study at any time, confidentiality and secrecy was fulfilled, research activities were not harmful, and the research was of benefit for the children and society as a whole.

Another specific ethical issue that we considered, one that applied to Studies I to III, concerned the use of data from the international AMPS
database located in Ft. Collins, Colorado, USA. When AMPS data on persons with different diagnoses or well persons is collected in any world region, the persons are asked for consent to submit their data to the database. However, for our studies, we accessed the data of the AMPS database without obtaining a new informed consent from the research participants related to our studies. Based on Brody’s argumentation (1998), we agreed that we could access the data of the database for our research purposes without informed consent because the following three conditions applied. First, we identified that consent was impractical because the data were collected over many years and from all over the world. It would have been too impractical and too expensive to seek secondary informed consent. Second, confidentiality was safeguarded, meaning that the data within the AMPS database cannot be traced back to research participants and therefore, anonymity is retained. Third, we think that our research was of significant importance for the research participants, occupational therapy practice, and society as a whole (Brody, 1998). We are aware, however, that one could question this decision because one important principle—the principle of autonomy (Swedish Medical Research Council, 1999)—was not applied. Neither informed consent of research participants or important others was obtained for the planned research with the consequence that the participants of Studies I to III had no opportunity to decide if they wanted to be participants, and if they wanted to withdraw at any time. Therefore, the ethics committees of the Umeå University and the Canton Zurich had to approve the research and had to weight the risks with the benefits of the planned research. Both ethical committees approved the studies.

Last, but not least, many of the issues that were mentioned above also apply to Study IV. Additional issues that needed attention were related to the design of the study. We implemented a feasibility study of a planned larger RCT with cross-over design to ensure that every child received the same occupational therapy intervention. Furthermore, the children and their parents were informed that they could withdrawal at any time of the study. Some of the parents declined to participate because they did not want to accept the waiting time. It was ensured that those children received the same occupational therapy intervention as did the children who participated in the study. In order to minimize outside influence and manipulation, we openly accounted for our purposes and methods, and registered Study IV at the International Standard Randomized Controlled Trial Number Register (ISRCTN).
RESULTS

Evidence Supporting Validity of the AMPS

The purpose of this project was to contribute evidence to support the valid use of the AMPS scales and measures in occupation-based and occupation-focused evaluation and intervention. I will, therefore, present the results of Studies I to IV in terms of validity evidence from different sources. Thus, in this frame, I will only present the results of Study IV that pertain to validity evidence that supports the use of the AMPS as an evaluation tool to identify a child’s strengths and limitations of ADL task performance, to guide the planning of occupation-based and occupation-focused interventions, and to evaluate the effectiveness of those interventions within the context of a planned larger RCT.

Evidence Related to Internal Structure

Stability of Item Difficulty Calibration Values (DIF)

In Study I, we found that overall the hierarchy of the 16 ADL motor and 20 ADL process items remained stable. Only one ADL motor item and none of ADL process items demonstrated a meaningful DIF contrast. More specifically, across the 112 comparisons (7 comparison pairs x 16 ADL motor items) for the ADL motor items, three differed by at least ±0.55 logit (2.68%) and all three were for the ADL motor item Aligns.

When we investigated further in an attempt to identify possible sources of DIF for the ADL motor item Aligns (i.e., if DIF was related to specific raters, age, gender, or version of the AMPS manual (English vs. German translation) (Fisher, 2011), we found that eight of 117 Middle European raters scored Aligns unexpectedly high compared to raters from the other world regions. Unexpectedly high ratings are very unusual for the ADL motor item Aligns as it is one of the easiest ADL motor items. Usually, when rater error occurs in relation to scoring Aligns, unexpectedly low, not high ratings are observed. Further investigation revealed no misfit of raters or any other evidence of rater scoring error among these eight raters. Interestingly, all eight raters who scored Aligns relatively high were calibrated as stricter than average raters when scoring the items on the ADL motor scale. This increases the likelihood they would score Aligns lower than expected, not higher. No other systematic patterns associated with participant age, gender, or version of the AMPS manual could be identified that could explain the DIF.
Evidence Based on Relation to Other Variables

*Stability of the Mean AMPS Measures of Typically-Developing Children from Middle Europe Compared to Mean AMPS Measures of Typically-Developing Children from Other World Regions*

In Study II, we evaluated if the mean ADL motor and mean ADL process ability measures among children from Middle Europe differed significantly compared to the means for children from each of the six other world regions, and if those differences where clinically meaningful. The results of the two way ANOVA for mean ADL motor ability measures revealed a significant effect for world region, $F=18.89$, $df=(6, 11091)$, $MSE=0.20$, $p<.001$, $\eta^2=.01$, a significant effect for age, $F=253.47$, $df=(13, 11091)$, $MSE=0.20$, $p<.001$, $\eta^2=.23$, and a non-significant region*age interaction effect, $F=1.20$, $df=(78, 11091)$, $MSE=0.20$, $p=.107$, $\eta^2=.008$. The results of the second two way ANOVA analysis for mean ADL process ability measures also revealed a significant effect for region, $F=15.32$, $df=(6, 11091)$, $MSE=0.17$, $p<.001$, $\eta^2=.008$, a significant effect for age, $F=292.13$, $df=(13, 11091)$, $MSE=0.17$, $p<.001$, $\eta^2=.26$, and a significant region*age interaction effect, $F=1.48$, $df=(78, 11091)$, $MSE=0.17$, $p=.004$, $\eta^2=.01$. In spite the presence of statistical significance, the magnitude of differences in mean ADL motor and mean ADL process ability measures were clinically not meaningful.

Given the significant region effect for mean ADL motor and ADL process ability and the significant region*age interaction effect for ADL process ability, we proceeded to estimate contrasts between Middle Europe and the other world regions. Out of 84 contrasts for ADL motor ability, four were statistically significant (2.38%) and for ADL process ability, three were statistically significant (1.79%). Furthermore, only two ADL process ability means for Middle Europe differed by more than $\pm 1$ SE from the international AMPS age-normative means; none exceeded $\pm 1$ SE for ADL motor ability.

*Sensitivity of the AMPS Measures for Discriminating Between Typically-Developing Children and Children With and at Risk for Mild Disabilities*

The results of the first forced regression analysis performed for Study III revealed no significant differences between groups in mean ADL motor ability measures, $B=0.021$, $CI=-0.061\text{ to }0.103$, $p=0.308$, $t=0.503$, $R^2=.361$. In contrast, there was a significant main effect for age $B=0.139$, $CI=0.129\text{ to }0.149$, $p<.001$, $t=26.187$, $R^2=.361$ as well as a significant age*group interaction effect, $B=-0.031$, $CI=-0.038\text{ to }-0.021$, $p<.001$, $t=-6.612$, $R^2=.361$. 

46
Results

The second forced regression analysis revealed significant differences between the two groups in mean ADL process ability measures, $B = -0.166$, $CI = -0.242$ to $-0.090$, $p < .001$, $t = -4.296$, $R^2 = .367$. Furthermore, there was a significant main effect for age, $B = 0.122$, $CI = 0.112$ to $0.132$, $p < .001$, $t = 24.751$, $R^2 = .367$ and a significant age*group interaction effect, $B = -0.027$, $CI = -0.035$ to $-0.019$, $p < .001$, $t = -6.204$, $R^2 = .367$.

Given the significant age by group interaction effects, we proceeded to perform post hoc $t$ test evaluations for significant group differences in ADL motor and ADL process ability at each age. With the exception of ADL motor ability measures for 4-year-olds, the children without a diagnosis in all age groups had significantly higher mean ADL motor and ADL process ability measures than did the children with mild disabilities ($t \geq 3.62$, $p < .001$). Effect size calculations revealed that most ADL motor effect sizes were medium to large beginning at age 7, and ADL process effect sizes were medium at 4 and 5 years and large beginning at 6 years.

Thus, we concluded that the AMPS ADL motor and ADL process measures are valid for discriminating between typically-developing children and children with and at risk for mild disabilities beyond the age of 4. More specifically, differences in mean ADL process ability between typically-developing children and children with and at risk for mild disabilities differed in an extent that was of crucial practical and clinical importance (i.e., $d$ was between 0.77 and 1.22 beyond the age of 4) (Cohen, 1988).

Evidence Related to Consequences of Testing and Test Fairness

**Differential Test Functioning (DTF)**

In Study I, a total of six comparisons (Middle Europe compared to six other world regions) for ADL motor ability measures and six comparisons for ADL process ability measures were made and none revealed any evidence of DTF. That is, all paired ADL ability measures of both ADL scales fell within the 95% control lines, and none of the paired measures differed by more than 0.09 logit.

**Use of the AMPS in the Context of a Feasibility Study**

In this section of the thesis, the results of Study IV are described from a validity perspective focusing on the use of the AMPS as an evaluation tool in a context of a feasibility study. I will, therefore, only describe aspects of the study that are directly related to evidence of validity as well as other aspects that had an impact on validity evidence for the AMPS. Our results
are reported according to main themes based on objectives of feasibility studies proposed by Thabane et al. (2010) and Tickle-Degnen (2013).

When evaluating the use of the AMPS for a full scale RCT with regard to process aspects, we identified the following strengths and weaknesses: Occupational therapist practitioners reported no problems with the amount of time for conducting the AMPS evaluation or with participants’ burden for data collection procedures. General problems that arose were more related to the lack of experience of occupational therapy practitioners in using standardized evaluation tools. Furthermore, not all occupational performance problems of the participating children were covered with the AMPS. Moreover, the results of the feasibility study indicated a) practitioner’s perceived problems and their need for ongoing discussion to support them in identifying and implementing ADL tasks of sufficient challenge for children, and b) technical issues and poor quality videotaping, resulting in the blinded AMPS rater reporting problems with rating all ADL motor and ADL process items based on the videos.

We assessed resources with regards to time, equipment, and material needed when using the AMPS as an evaluation tool in an intervention study. Our original time schedule needed to be extended from 12 to 20 months. The main reason being that we underestimated the time needed for a) initial training and on-going communication with the occupational therapy practitioners related to conducting research and the importance of adhering strictly to the randomization process, and b) communication with the pediatricians related to the recruitment of participants. While the occupational therapy practitioners were highly motivated and engaged in the training required as a basis for participating in the research project, the training was not sufficient to address the complexity inherent in implementing the study in a context where research in occupational therapy was just emerging. In relation to the use of the AMPS, many discussions of the standardized administration procedures and the documentation required by the study protocol were necessary. Consequently, some participants were “lost” from participation and the time needed lengthened. In contrast, the occupational therapists did report that their participation in Study IV resulted in their ability to improve their evidence-based practice (EBP) skills in general, and in relation to the use of standardized evaluation tools specifically. They also reported no problems concerning space, equipment (e.g., video camera), and materials (e.g., groceries for AMPS tasks). Only the unforeseen technical problems with videotaping of the AMPS observations for research purposes required additional time, knowledge, and skills as well as additional resources (e.g., IT support).

When we assessed management in relation to documentation according to the study protocol, we focused on issues that had the potential to impact on the valid use of the AMPS in the context of an intervention study. We found that data collection was associated with the reliable management of timely AMPS observations and reliable blinded AMPS
Results

rating. No challenges were met with data entry or storage. Despite the challenges of the occupational therapy practitioners concerning the use of standardized evaluation tools, they were able to plan interventions based on the results of the AMPS and COPM evaluations and carry out the interventions with full adherence to the study protocol. They reported that using the AMPS (and the COPM) were helpful in that process.

We assessed the scientific basis of the feasibility study in relation to the use of the AMPS by evaluating the criteria for clinical meaningfulness in relation to observed clinically meaningful differences among the participants. More specifically, a clinically meaningful change between two AMPS measures is 0.3 logits on both scales (Fisher & Jones, 2012a). Out of 17 participants, 9 (52.9%) demonstrated increases of at least 0.3 logits on at least one of the AMPS scales during the intervention. We did not find any patterns among those who did or did not improve in quality of ADL task performance related to diagnoses, gender, age, or group allocation.
DISCUSSION

Discussion of Evidence Supporting Validity of the AMPS

Evidence Related to Internal Structure

*Stability of Item Difficulty Calibration Values (DIF)*

Study I was a validity study based on evidence to support the internal structure of the AMPS scales, specifically the stability of the item difficulty calibration values of the AMPS (DIF). Our results are congruent with studies of DIF finding that overall, the item difficulty calibrations values of the AMPS remained stable across world regions (Fisher & Merritt, 2012b; Goldman & Fisher, 1997; Magalhães et al., 1996). One aspect we found interesting is that our results did not confirm DIF on the ADL motor item *Endures* (Fisher & Merritt, 2012b), but instead revealed DIF on the ADL motor item *Aligns*.

Nevertheless, there is general agreement that consideration of DIF is of crucial importance for sound use of evaluation tools in practice and research (AERA et al., 1999). DIF occurs when persons of approximately equal ability differ in their responses to an item according their group membership (e.g., persons from Middle Europe compared to persons from North America) (AERA et al., 1999). In Study I, the ADL motor item *Aligns* demonstrated DIF in three comparisons between Middle Europe and the other regional groups, but did not demonstrate DIF when Middle Europe was compared to the item difficulty calibration values based on the total sample (see Table 6).

Because of the importance of DIF, we explored our data further to determine if we could identify possible sources of DIF. More specifically, we explored if there were actual differences in item responses based on group membership (e.g., unequal proportions of more vs. less able persons among regional groups), if there was identifiable rater scoring error, and/or if sample size was the source of the identified DIF.

The ADL motor item *Aligns* pertains to performance of ADL tasks without the persistent need for propping (Fisher & Jones, 2012b). The higher item calibrations values for *Aligns* for the Middle European sample (see Table 6) indicated that persons from Middle Europe were rated higher on the ADL motor item *Aligns* (i.e., had lower need to persistently prop) than were persons from other world regions. In other words, overall the persons from Middle Europe had very few problems with *Aligns*, meaning that they were only very rarely observed to persistently prop during their ADL task performances.
Discussion

When we considered quality of ADL task performance among different “diagnostic” groups (i.e., those groups that were more likely to have higher scores on Aligns vs. those who were more likely to have lower scores), we considered that well, healthy persons and persons at risk for or who have been diagnosed with mild disabilities are two groups that are at least likely to need to prop; hence, they likely will have higher scores. Those persons who are most likely to prop are persons with more severe disabilities (e.g., spinal cord injury, multiple sclerosis, brain injury) (Fisher & Merritt, 2012b). When we compared the samples, the Middle European sample overall included a higher proportion of persons with mild disabilities, which could have resulted in Aligns having a higher item difficulty calibration value for the Middle European sample. However, the Middle European sample also included a lower proportion of well persons, which would likely be associated with a lower, not higher calibration value for Aligns in the Middle European sample compared to the other regions. Therefore, we concluded that different proportions of persons with higher versus lower ADL ability was not the basis of the cross-regional DIF for Aligns between Middle Europe and other world regions.

Table 6
Item Calibration Values and DIF Contrasts (in Logits) of the ADL Motor Skill Item Aligns of the Middle European Sample Compared to Samples From Other World Regions

<table>
<thead>
<tr>
<th></th>
<th>Other world region</th>
<th>Middle Europe</th>
<th>DIF contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>North America</td>
<td>0.26</td>
<td>0.91</td>
<td>0.65*</td>
</tr>
<tr>
<td>United Kingdom/Republic of Ireland</td>
<td>0.27</td>
<td>0.91</td>
<td>0.64*</td>
</tr>
<tr>
<td>Nordic Countries</td>
<td>0.50</td>
<td>0.91</td>
<td>0.41</td>
</tr>
<tr>
<td>Other Europe</td>
<td>0.56</td>
<td>0.91</td>
<td>0.35</td>
</tr>
<tr>
<td>Australia/New Zealand</td>
<td>0.33</td>
<td>0.91</td>
<td>0.58*</td>
</tr>
<tr>
<td>Asia</td>
<td>0.37</td>
<td>0.91</td>
<td>0.54</td>
</tr>
<tr>
<td>Total sample</td>
<td>0.53</td>
<td>0.91</td>
<td>0.38</td>
</tr>
</tbody>
</table>

Note 1: The higher the item difficulty calibration value, the easier the item.
Note 2:* represents contrasts that were ≥0.55 logits that we used as criteria of significant DIF.
Nevertheless, we do acknowledge that diagnosis can be a source of DIF. That is, the results of earlier studies have shown that the AMPS items do display diagnostic DIF. More specifically, when Fisher and Merritt (2012b) investigated cross-diagnostic DIF, 22.2% of the paired comparisons of the ADL motor items and 5.4% of the paired comparison of the ADL process items revealed DIF. Interestingly, the ADL motor item Aligns and Moves where the only two motor items that displayed DIF in less than 5% of the time. That evidence supports our conclusion that variations in our samples associated with diagnosis and/or high versus lower ADL ability likely was not the source of the DIF on the ADL motor item Aligns. Obviously, our evaluation of “diagnostic” causes for DIF were only preliminary, and thus, we recommend further investigation.

Secondly, we considered rating scoring error as a possible source of DIF. Clearly, the Middle European raters had scored Aligns higher compared to raters to other world regions. Given that it seems unlikely that this was due to sample characteristics (i.e., high vs. low ADL ability), we considered the possibility that linguistic issues related to which translation of the AMPS manual was used might have been a source of rater scoring error that resulted in DIF. We found, however, no difference between raters who had scored using the original English manual and those who were likely to have used the German translation. We did, however, identify eight Middle European raters who were particularly prone to rate Aligns higher than expected. To determine if their ratings were the source of the DIF, we removed these eight Middle European error raters from the analyses, which minimized the magnitude of DIF, but did not resolve it. Hence, we concluded that the source of the DIF could not be attributed to a few raters. If rater scoring error was the issue, it likely was a more “broad based,” and perhaps related to culture. For example, Goto et al. (1996) found differences in rater severity between Japanese and European raters. European raters tended to be stricter when they rated Japanese than when they rated Europeans. Yet, overall, it is likely that within our regional samples, those from Middle Europe rated primarily Middle Europeans, and those from, for example, North America, rated primarily North Americans; all regional sample may have included persons who immigrated to those regions, but data on the regional origin of the participants was not available.

Yet another possibility we considered was that the error was not among the Middle European raters, but instead among raters from the other world regions. To clarify, as I noted earlier, Aligns is among the easiest of the ADL motor items of the AMPS. This means that it is expected to be given higher overall ratings. Yet, during AMPS rater training courses, it is very common for course participants to want to score Aligns lower, especially among persons with neurological conditions, because the course participants suspect that low tone affects the person’s ability to bear weight evenly on both legs. In this case, the participants are clearly demonstrating rater scoring error because a) the participants are scoring a body function (e.g., tone), not occupational performance, and b) they
are not scoring observable behavior (A. G. Fisher, personal communication, March 30, 2013) Thus, it remains possible that the Middle European raters are among the only group of raters who score the ADL motor item *Aligns* in a valid manner, but further research would be needed to support or refute this possibility.

Last, but not least, we considered the sample size as a possible source of DIF (AERA et al., 1999). Our regional sample sizes was large, and large sample sizes result in too much power in relation to p values. Our data analyses revealed p values lower than 0.001 in all 252 comparisons. We tried to avoid over-identifying significant differences based on p values by acknowledging *a priori* that we would have too much power, and by using effect sizes (not p values) as our method of identifying DIF (Tristán, 2006; Wilson, 2005). Thus, large sample sizes do not explain our findings of significant DIF for the ADL motor item *Aligns* because Tristán’s method is considered to be robust against sample size (Tristán, 2006).

While we were unable to identify a clear reason for the identified DIF for the ADL motor item *Aligns*, this is not unusual. Other author’s report difficulties in identifying the causes or in understanding the sources of DIF (AERA et al., 1999; Osterlind & Everson, 2009). It is also possible that the DIF we identified is due to random error as *Aligns* has not been shown in earlier studies to demonstrate DIF (Fisher & Merritt, 2012b). Moreover, because the item difficulty calibration values of the AMPS items remained overall stable among groups, we concluded that the results of Study I supported existing validity evidence in relation to the internal structure of the AMPS scales. Thus, the results support the validity of the AMPS as an occupation-based and occupation-focused evaluation tool that can be used in future cross-regional research and practice.

Yet, our results also suggest that further DIF studies are needed. Such studies might evaluate the internal structure of the AMPS scales across other world regions (e.g., Asia compared to other world regions) or across countries within one world region (e.g., Austria compared to Switzerland). Such studies should consider evaluating whether or not the ADL motor item *Aligns (or Endures)* continues to display DIF when other groups of persons are compared (e.g., Asia and North America).
Evidence Based on Relation to Other Variables

We evaluated two sources of evidence to support the validity of the AMPS measures based on relation to other variables. In Study II, we evaluated for evidence based on the stability of the mean AMPS measures of typically-developing children between Middle Europe and other world regions. Thus, region was the external variable. In Study III, we examined for evidence in relation to "diagnostic groups"—whether the AMPS measures were sensitive enough to discriminate between typically-developing children and children with and at risk for mild disabilities. In this case, diagnostic group was the external variable.

Stability of the Mean AMPS Measures of Typically-Developing Children from Middle Europe Compared to Mean AMPS Measures of Typically-Developing Children from Other World Regions

The results of Study II suggest that, overall, the mean AMPS measures of typically-developing children from Middle Europe compared to mean AMPS measures of typically-developing children from other world regions remained stable. More specifically, while the results of our ANOVA analyses revealed significant effects for mean ADL motor and ADL process ability by world region, the results of our contrast estimations revealed that 95.9% of the mean ADL motor and ADL process ability measures remained stable across world regions. These findings are consistent with the results of earlier research (Peny-Dahlstrand et al., 2012). The results of our analyses should be interpreted with caution because the estimated contrasts for the age groups 2-4 and 10-15 were based on sample sizes below 30 per cell. The result was large confidence intervals (CIs) and the risk of not enough power to identify significant differences at those ages. We did, however, take the small sample sizes into consideration by estimating contrasts based on the ANOVA model, implementing simultaneous CI procedures, and using Sidak corrections.

When we compared the differences between the Middle European means and the international age-normative means of the AMPS, we found that only two of the contrasts were larger than the mean SE, and both were for ADL process ability measures. Given this low number of the clinically meaningful contrast, we concluded that our results provided preliminary evidence that the international age-normative means of the AMPS (and their related standard deviations, SDs) are likely appropriate for use with children from Middle Europe. Our results are in line with Peny-Dahlstrand and colleagues (2012) who found that none of the mean differences in ADL motor and ADL process ability measures between children from North America and Nordic Countries or between the Nordic Countries and the international age-normative means of the AMPS exceeded a similar criterion of clinical meaningfulness.
However, it is also critical to note that for the age groups 5 and 15, the mean ADL process ability measures of children from Middle Europe were higher than were the means for those of the international age-normative means of the AMPS. This could suggest that comparison of the results of Middle European children to the international age-normative mean values of the AMPS might result in misinterpretations of Middle Europeans children’s overall ADL process ability. That is, children from Middle Europe may have a higher possibility of having ADL motor or ADL process ability measures that indicate more efficient ADL task performance than if they were to be compared to their age-matched Middle European typically-developing peers.

This idea that children from Middle Europe may be identified as having higher ADL ability when their AMPS measures are compared to the international age-normative means could also be considered as validity evidence related to test fairness and the absence of test bias (AERA et al., 1999; Camilli, 2006). Thus, the evidence from Study II related to validity in relation to other variables also has potential implications for evidence related to consequences of testing, but I chose to present it here because the relation to other variables was the primary focus of Study II.

While our results generally provide evidence that supports existing evidence (Peny-Dahlstrand et al., 2012) that the mean ADL motor and ADL process ability measures of the AMPS remain stable across world regions, there is clearly a need for continued research. Such studies should focus on evaluating differences in age-normative means of the AMPS by world region based on larger sample sizes. There is also a need to examine the cross-regional stability of the ADL motor and ADL process age-normative means for healthy, well adults.

**Sensitivity of the AMPS Measures for Discriminating Between Typically-Developing Children and Children With and at Risk for Mild Disabilities**

Considered together with results of earlier studies, the results of Study III supported the sensitivity of the AMPS for detecting differences in ADL ability among diagnostic groups who are likely to differ in mean ADL ability (cf. Merritt, Gahagan, & Kottorp, 2013; Oakley, Duran, Fisher, & Merritt, 2003; Parks, Rasch, Mansky, & Oakley, 2009; Peny-Dahlstrand et al., 2009; Poole, Atanasoff, Pelsor, & Sibbitt, 2006). More specifically, the results of Study III supported the sensitivity of the AMPS measures for discriminating between typically-developing children and children with and at risk for mild disabilities beyond the age of 4. Our results suggest that the ADL process measures of the AMPS are overall more sensitive than are the ADL motor measures because significant differences were identified at all ages, 4 to 15, and more importantly, the ADL process effect sizes were consistently moderate to large at all ages.
This latter finding provides an index of the standardized group difference. In contrast to \( t \) values, effect sizes are independent of sample size. That is, while there might be concern that we had too much power in our data and over-identified group differences, the use of Cohen’s \( d \) minimized the risk for too much power being an issue.

The large effect sizes found for ADL process ability as early as 6 years indicated that the mean differences between typically-developing children and a group of children that would be expected to have some problems overall with ADL task performance but fewer problems than would a group of children with more severe disabilities, further supported the sensitivity of the ADL process ability measures. Large effect sizes are of crucial practical and clinical importance (Cohen, 1988). We suspected that the large effect sizes emerged because children with DCD and ADHD are typically diagnosed by age 5 (APA, 2013; Blank, Smits-Engelsman, Polatajko, & Wilson, 2012), and thus, there was a higher probability that the children aged 5 and above in our sample actually differed from their age-matched, typically-developing peers. Reduced quality of and persistent problems in occupational performance (e.g., ADL or school work performance) are part of the diagnostic criteria for ADHD, DCD, and LD (APA, 2013).

The results of Study III also support the use of the AMPS as a sensitive and valid tool that can be used for occupation-based and occupation-focused evaluation in children with mild disabilities in occupational therapy practice and research. More specifically, while this was not a direct focus of our study, by implication, we feel that our results suggest that the AMPS is a valid tool that can be used to evaluate ADL performance strengths and weaknesses of children with mild disabilities and to use the results of the AMPS observation as a basis for designing and implementing occupation-based and/or occupation-focused interventions intended to increase children’s ADL abilities.

Moreover, while I chose to discuss the ability of the AMPS measures to discriminate between groups under validity evidence related to other variables, our results also contributed to evidence related to consequences of testing. That is our results provided evidence of the ADL task performance problems of children with and at risk for mild disabilities. These results might influence decisions about the use of the AMPS in pediatric occupational therapy practice and subsequently influence decisions about the need for and benefits of occupational therapy services. For example, future research might focus on investigating the sensitivity of the ADL motor and ADL process ability measures of the AMPS for detecting change following an occupational therapy intervention in children with mild disabilities. As I will discuss in more detail below, we obtained only very preliminary evidence related to sensitivity of the AMPS measures for evaluating change following intervention in our feasibility study.
Evidence Related to Consequences of Testing and Test Fairness

Differential Test Functioning (DTF)

As the presence of DIF can signal a risk to the internal structure of a scale that, in turn, has consequences related to test bias, we evaluated further to determine if the DIF present in the AMPS items Aligns results in DTF. Our results revealed no DTF, and hence, no evidence of test bias. Considered together with results of earlier studies (e.g., Fisher & Merritt, 2012b), the results of Study I suggest that the ADL motor and ADL process measures are free of cross-regional DTF. Despite the differences in item difficulty calibration values among regional groups (Goldman & Fisher, 1997; Magalhães et al., 1996), there is no evidence of systematic test bias (DTF) of the AMPS measures in relation to world region (Fisher & Merritt, 2012b). Therefore we assume, that when the AMPS is used to evaluate ADL performance of persons from different world regions, the results allow for fair comparisons and do not favor one regional group compared to another, as long as AMPS raters take their and the person’s cultural background into account (Clawson, 1995; Goto, Fisher, & Mayberry, 1996).

Use of the AMPS in the Context of a Feasibility Study

In this thesis, the results of Study IV are discussed from a validity perspective that may contribute to a successful planning and implementation of further research, specifically a larger full-scale RCT.

First, our evaluation of the use of the AMPS with regard to process, resources, and management aspects of the feasibility study revealed that there was no perceived burden among the occupational therapy practitioners who participated in the study that was related to time, equipment, or materials (e.g., groceries needed for AMPS observations). This lack of burden extended to the use of the AMPS software by the blinded rater for data entry and generation of each child’s ADL motor and ADL process ability measures.

In contrast, there were problems with the use of standardized evaluation tools. Among the issues identified were a) the lack of experience of the occupational therapy practitioners with using standardized tests, and b) the resultant need for many discussions related to the importance of adhering to standardized evaluation procedures. While the discussions extended beyond the use and administration of the AMPS to the overall need for adherence to the standardized administration procedures of other evaluation tools (e.g., the use of the COPM [Law et al., 2009]), they did include issues related to the use of the AMPS. For example, there was an ongoing concern that the occupational therapy practitioners could not find tasks that were of sufficient challenge. Yet, analysis of their data
Discussion

revealed no evidence that the AMPS tasks the children performed were not of sufficient challenge. That is, allowing a child to perform an AMPS task that is too easy results in the child’s ADL motor and/or ADL process ability measure to be unexpectedly high, and such error was never identified. Therefore, we concluded, that while the occupational therapist practitioners perceived problems, they did ensure that the children performed ADL tasks of appropriate challenge.

Our results concerning the occupational therapy practitioners’ problems with following standardized evaluation procedures are in line with results of earlier studies reporting similar problems in commitment and adherence of therapists to following standardized evaluation procedures (Fange, Risser, & Iwarsson, 2007; Kjellberg, Bolic, & Haglund, 2012). The reasons have varied, for example, general resistance to cooperation (Fange et al., 2007), level of knowledge, and lack of time and experience (Chard, 2000, 2004; Kjellberg et al., 2012). Similarly, earlier studies reported difficulties among occupational therapists to administer the AMPS using standardized administration procedures (e.g., Chard, 2000, 2004; Hitch, 2007).

In contrast to their expressed concerns and experienced problems, the occupational therapist practitioners also reported that their participation in the feasibility study, including the use of the AMPS, resulted in increased awareness and improvements in their practice from an evidence-based perspective. These results are supported by Thomas and Law (2013), who conducted a literature review and found that participation in research, adequate support, and research training was an important factor that supported the use of evidence-based practice (Thomas & Law, 2013). From a validity perspective, these results imply that the participation of occupational therapy practitioners in research might increase their knowledge of and skills in the valid use of standardized evaluation tools specifically and their evidence-based practice skills in general.

Furthermore, a common issue raised by the occupational therapy practitioners was that the AMPS tasks were only related to ADL. Yet many of the children (or their parents) reported that the children’s problems included performance of schoolwork tasks, play, and behavior in the classroom or at home. The AMPS, however, is a test of quality of ADL task performance (Fisher & Jones, 2012a, 2012b). Thus, our findings suggested that there was a need to include other evaluation tools (e.g., the school version of the AMPS [Fisher et al., 2007]) so as to be able to evaluate a wider range of occupational performance problems among children with and at risk for mild disabilities.

Another issue specific to our feasibility study was that we used videotaping to enable the AMPS observations to be rated by a blinded rater. The occupational therapy practitioners experienced some sense of burden related to the use of video-equipment for videotaping the AMPS
observations for later scoring by the blinded rater, including the need for technical support. Furthermore, the blinded rater reported that she could not always see in the videotape what she needed to be able to observe to be able to rate some of the AMPS items. Yet, analysis of her data revealed that she did rate the AMPS in a reliable manner. It is also important to note that videotaping is never recommended when the AMPS is used in clinical practice (Fisher & Jones, 2012a). Therefore, we concluded that video-rating may impact the valid use of the AMPS for research purposes and, in turn, the valid interpretation of the mean ADL motor and ADL process ability measures.

A critical basis for generating comparable results and ensuring their validity is the accurate use of standardized test administration, scoring, and reporting procedures (AERA et al., 1999). AERA et al. (1999) suggested that evaluation procedures should carefully be carried out in a standardized way in order to ensure that persons, whose performance is being evaluated, receive comparable and equitable treatment during the whole evaluation process. When we interpret the results of Study IV in relation to test fairness, evaluated in the context where the evaluation tool was used (Camilli, 2006)—namely in a feasibility study—we conclude that “fairness as equitable treatment in the testing process” could not be established (AERA et al., 1999, p. 74). We concluded, therefore, that adherence to standardized evaluation procedures needs to be insured by practitioners and researchers when a full-scale RCT is implemented. The challenge of such a study will to ensure that the raters are blind to the phase of the study, while the participants (children and occupational therapists) are not burdened or otherwise subjected to “unfair” conditions.

In general, evaluation tools are administered in expectation that some beneficial information will be gained from the intended results (AERA et al., 1999). In occupational therapy practice and research, the information gained from results of an evaluation tool usually informs decisions, such as a person’s need for occupational therapy services or evidence of effectiveness of interventions. In terms of validity evidence in relation to consequences of testing, it is, therefore, important to indicate if the benefits of an evaluation are likely to be realized (AERA et al., 1999). The results of our feasibility study revealed that the AMPS measures are likely sensitive enough to detect positive benefits from an occupational therapy intervention designed to improve ADL task performance. That is, as I noted in the Introduction, feasibility studies do not include the use of inferential statistics to evaluate for statistical significance (Leon et al., 2011; Tickle-Degnen, 2013). Yet, our preliminary results of the evaluation for clinical meaningful changes did suggest that the AMPS measures will be sufficiently sensitive to detect changes in our planned full-scale RCT.
Implications and Recommendations for Occupational Therapy Practice and Research

The main focus in this section is how the further validity evidence gained in this PhD thesis may contribute to the use of the AMPS in occupation-based and occupation-focused evaluation and intervention in occupational therapy practice and research.

Implications for Occupational Therapy Practice

- The overall absence of DIF and the equivalence of the age-means for typically-developing children across world regions suggests that the AMPS is valid for use in clinical practice not only in Middle Europe and with children, but also in other world regions. This suggests that occupational therapy practitioners internationally can use the AMPS to evaluate quality of ADL task performance in children in order to plan and evaluate the effectiveness of occupation-based and occupation-focused interventions.

- The international age-normative mean values of the ADL motor and ADL process ability measures can preliminarily be applied to children from Middle Europe and, therefore, allow occupational therapy practitioners from Middle Europe to make valid norm-referenced interpretations of AMPS observation results based on the international normative values.

- The AMPS ADL motor and ADL process ability measures are valid for purposes of evaluating and identifying ADL problems among children with and at risk for mild disabilities older than age 4.

- The detected problems in ADL task performance of children with and at risk for mild disabilities are of crucial practical importance and support the need to evaluate for ADL performance problems among these children, and if needed, to design and implement effective occupation-based and occupation-focused interventions that increase children’s ADL abilities.
Discussion

Implications and Recommendations for Further Research

- The overall absence of DIF and the equivalence of the age-means for typically-developing children across world regions suggests that the AMPS is valid for use in research not only in Middle Europe and with children, but also in other world regions. This suggests that the AMPS can be used in international research as a valid occupation-based and occupation-focused tool to evaluate quality of ADL task performance.

- The AMPS ADL motor and ADL process measures are valid for research purposes to evaluate the quality of ADL task performance of children with and at risk for mild disabilities older than age 4.

- When researchers use the AMPS for research purposes they need to ensure that the AMPS is administered, scored, and reported according to the standardized administration procedures. For researchers in the context of Middle Europe, where research in occupational therapy is just emerging, it is important to plan enough time for training and ongoing supervision of practitioners, according to their knowledge and skills in evidence-based practice in general and in the use of standardized evaluation tools specifically.

- When the AMPS is used in research, it is important to recognize that the AMPS is an evaluation tool of quality of ADL task performance, and that it is not sufficient to address other types of occupational performance problems that are not covered with the AMPS. Thus, if the targeted areas of concern pertain to schoolwork or behavior, other occupation-based and/or occupation-focused outcome measures will be needed. Consideration of what evaluation tools to use should always take participant burden into consideration. The choice of evaluation tools should also take into consideration the sensitivity of the measures and sensitivity might be more important to consider than time burden.

- We recommend further DIF studies that might evaluate the stability of item difficulty calibration values of the AMPS across other world regions or across countries within single world regions. Such studies should take into consideration evaluation of whether or not the ADL motor item Aligns continues to display DIF and if so, try to identify possible sources of DIF.

- Furthermore, we recommend verifying the stability of the mean ADL motor and the mean ADL process measures of the AMPS of typically-developing children from Middle Europe and from other world regions with a larger sample, namely with 50 to 70 children per cell.
• There is also a need to examine the stability of the mean ADL motor and the mean ADL process measures for the adult normative sample of the AMPS among world regions.

• Moreover, our results suggest that the AMPS is a valid evaluation tool that can be used to evaluate ADL performance strengths and weaknesses of children with mild disabilities. Further research could evaluate how the results of the AMPS observation (e.g., the clusters of ADL skill strengths and problems) can be used as a basis for designing and implementing occupation-based and/or occupation-focused interventions intended to increase children’s ADL abilities.

• Our results of the evaluation of clinical meaningfulness of change provided preliminary evidence that the AMPS will be sufficiently sensitive to detect change in full-scale RCTs. However, further research needs to evaluate the sensitivity of change of the ADL motor and ADL process ability measures of the AMPS with children with and at risk for mild disabilities.
Other Methodological Considerations

The main focuses in this section of the discussion are methodological considerations that have not been raised before.

Other Methodological Considerations Related to Effect Sizes

In Studies I to III, the overall sample sizes were very large, resulting in a risk for too much power and over-identifying significant differences based on *p* values alone (Cumming, 2012). Therefore, we used effect sizes to describe the magnitude and meaning of differences among the groups from a clinical perspective.

In Study I we used effect sizes to evaluate for significant DIF because *p* values are not recommended as the sole criteria for determining DIF. (Tristán, 2006; Wilson, 2005). More specifically, we set our criteria for the presence of significant DIF based on a logit difference of at least ±0.55 logit (Tristán, 2006). This criterion could be questioned as there is little agreement about the criterion for when a contrast signals DIF. Recommended and used values vary between 0.40 and 0.64 logit (Conrad, Dennis, Bezruyczko, Funk, & Riley, 2007; Draba, 1977; Linacre, 1994; Tennant & Pallant, 2007; Tristán, 2006; Wilson, 2005) and the most commonly used is 0.50 logits (e.g., Fisher & Merritt, 2012b). However we decided to use the criterion of ±0.55 logits based on Tristán’s (2006) thorough analyses (cf. Introduction). This criteria was used in an earlier cross-regional study that evaluated differences in item difficulty calibration values of the School Version of the AMPS (Munkholm et al., 2010).

In Study II, we used the mean *SE* of the ADL motor and ADL process ability measures of the AMPS as criteria for determining meaningful differences between the regional groups. More specifically, we set our criteria for the presence of meaningful differences as a logit difference of 0.25 for mean ADL motor ability and 0.20 for mean ADL process ability measures. We decided to use the mean *SEs* of the ability measures of the AMPS because they were more stable than the *SDs* of our estimated contrasts based on the broad variation of participants per cell. We had considered common effect sizes measures (e.g., Cohen’s *d* or Hedges’ *g*) for evaluating standardized group differences, but decided not to use them as their calculations are based on raw contrasts, not estimated contrasts (Cohen, 1988).

The use of the mean *SEs* of the ADL ability measures of the AMPS as criteria for determining meaningful differences among groups could be questioned, but our criteria were stricter than other criteria that have been used in earlier studies. For example, when important differences in
age-related means for basic ADLs were compared for the PEDI, SDs were used (Berg, Aamodt, Stanghelle, Krumlinde-Sundholm, & Hussain, 2008; Groleger, Vidmar, & Zupan, 2005; Haley et al., 1992); the use of the SDs is a common and well-established criterion to describe standardized group differences (Cohen, 1988; Cumming, 2012; Ellis, 2010). Considered together—varying sample sizes per cell and the use of estimated contrasts based on the linear model (instead of raw contrasts)—we decided to use the mean SEs of the ADL ability measures of the AMPS as criteria for interpreting clinical meaningful differences. These criteria have been implemented in prior AMPS studies (Fisher & Merritt, 2012a; Sperens et al., 2012).

In Study III we used Cohen’s d as an index of how much two groups differed when compared with their pooled SDs (Cohen, 1988). Because the groups of typically-developing children and children with and at risk for mild disabilities were of dissimilar size, we could have implemented Hedges’ g instead, where each group’s SD is weighted by its sample size (Hedges, 1981 in Ellis, 2010). However, we decided for Cohen’s d because the SDs of the groups did not differ and the homogeneity of variance assumption was not violated (Ellis, 2010).

Other Methodological Considerations Related to Study IV

In Study IV we analysed our data following deductive content analysis approach (Mayring, 2010), using the objectives of feasibility studies as main categories (Thabane et al., 2010; Tickle-Degnen, 2013). Given that qualitative research usually is inductive, meaning that categories emerge from the data (Schreier, 2012), one could question our approach. Furthermore, we did not involve children and/or their parents in the evaluation of feasibility, meaning that their views did not contribute to refining the study. Moreover, the research team evaluated feasibility from the researcher’s perspective, using peer debriefing with other researchers and member checking with occupational therapy practitioners for establishing credibility (Polit & Beck, 2008). However, an evaluation of feasibility from an outsider perspective might be more objective and might reveal other aspects of feasibility.
CONCLUSIONS

The results of Studies I to IV provided evidence to support the validity of the AMPS scales and measures for use as an occupation-based and occupation-focused evaluation tool.

- The DIF analyses showed that overall the item difficulty calibration values of the Middle European sample remained stable compared to samples from other world regions. Therefore, our results provided evidence to support the validity of the internal structure of the AMPS scales across the world regions included in this study.

- The DTF analyses revealed no evidence of systematic test bias of the AMPS in relation to world region and support the use of the AMPS in occupation-based and occupation-focused evaluation with persons in different world regions.

- Evaluation of evidence to support validity in relation to the use of the AMPS with typically-developing children across world regions suggested that, overall, the mean AMPS measures of typically-developing children from Middle Europe remained stable compared to the mean AMPS measures of typically-developing children from other world regions. Therefore, the international age-normative means of the AMPS are likely to be applicable to children from Middle Europe.

- Evaluation of evidence to support the validity of the AMPS in relation to the sensitivity of the AMPS measures for discriminating between typically-developing children and children with and at risk for mild disabilities revealed that the AMPS measures likely can be used to a) identify those children who have problems with ADL task performance and are in need for occupational therapy services, and b) provide valid ADL performance measures in occupation-based and occupation-focused evaluations of children with and at risk for mild disabilities.

- When occupation-based and occupation-focused evaluations were used in the context of a feasibility study, and the validity evidence for the use of the AMPS as a standardized evaluation tool was evaluated, our results indicated that a) the use of the AMPS is feasible in relation to needed time, equipment, and material; b) the occupational therapist’s implementation of the standardized administration procedures of the AMPS need to be supported; and c) the AMPS measures will likely be sufficiently sensitive to detect changes in quality of ADL performance in full-scaled RCTs.
ACKNOWLEDGEMENTS

A few years ago, I started this long trip. It was like tramping in the mountains; like enjoying colorful meadows; like stepping over pebbles; and yet, constantly heading towards the goal. On my way, different people came along with me, the entire way, or parts of it. I was blessed with enriching encounters that helped me to develop on a personal and professional level. Now, I have finished. Therefore, I would like to thank:

All the participants in Middle Europe and around the world, who participated in my studies and all occupational therapists in Austria, Germany, Slovenia, and Switzerland, who collected the data for my studies. Especially, the occupational therapists in Switzerland for being brave and breaking new ground by participating in an intervention study.

Anne G. Fisher, my main supervisor through most of my PhD studies, for sharing your knowledge generously, for teaching me thoroughly, for challenging me in a friendly way, for strengthening my confidence in my work and for giving me warm support. Thanks for your great effort and engagement. Your wisdom and your ability to stay down-to-earth have always impressed me.

Ingeborg Nilsson, who took over the role of the main supervisor for the last year of my PhD studies, for all your professional and personal support. I learned from you not only how to design and implement research into practice, but how to overcome barriers by thorough problem solving and relying on a huge portion of humor. Thanks a lot for your warm introduction into the Swedish culture.

Julie Page, my Swiss supervisor, for taking up my ideas, for supporting me to transform them into a grant application, and for admitting me into your research team. Thank you, for exploring with me the Swiss jungle of occupational therapy research and for supporting me to find my confidence in the role as researcher. Your ability to express critical issues in a nice way is unbeatable.

Birgitta Bernspång, the head of the Division of Occupational Therapy in Umeå, for welcoming me so warmly at your division, for inviting me to your house, and for providing bikes during my stays. I enjoyed roaring with laughter with you in different situations.

André Meichtry, co-author and colleague, for your generous statistical support, your great engagement, and your patience. You have invited me to dig deeper and strongly facilitated my growth in the field of statistics.

Berg Brett, for your excellent, immediate, and kind support, and provision of diagrams and data files from the AMPS database.
Acknowledgements

Christoph Künzle, co-author, for your great effort supporting the planning and implementation of the intervention study and for your engagement in strengthening occupational therapy in Switzerland. Your knowledge and engagement with children with disabilities inspire one to work with these clients, either in practice or by doing research.

Sabine Hendriks, Ursula Gubler-Thomann, and Christiane Mentrup for your openness and encouragement in the beginning of this PhD thesis. Your belief in my ideas enabled starting this trip.

My PhD colleagues in Umeå for sharing and discussing with me research ideas, methods and the implementation of research, for going out with me, for inviting me into your family homes, for taking me to the spa and to soccer plays, and for just having fun. To Magnus Zingmark and Ellinor Larsson, for valuable discussions about effect sizes, the implementation of interventions studies, and data analyses; and most notable, for enjoying fika on social-internet-based media.

My colleagues at the Institute of Occupational Therapy in Winterthur, for your interest in my work and for your encouragement. You probably don’t know how much your earnest interest motivated me to keep on walking. Thank you for your kind e-mails to the north, and for the music and the films that you packed into my suitcase that helped me to get through long stays away from home.

My PhD colleagues at the School of Health Professions in Winterthur, for sharing and bearing with me the experience of being a PhD student in a “new context.” I am so proud of you and I hope that your pioneering work has smoothed the way for successors.

Barbara Aegler and Daniela Senn for studying with me the AMPS and the OTIPM, for valuable and enjoyable collaboration, for interesting discussions about skill items and occupation-based and/or occupation-focused practice, for demonstrating competent ADL tasks performance (e.g., “Fotzelschnitte”), and for encouraging me from the first idea till the end of this PhD thesis.

My family and friends, who stayed by my side, for giving me encouragement and for supporting me to keep going.

Michael, Günter, Gisela, Franziska, Edwin, and Cornelia for always giving me your support, for your incredible patience, and for all your love. I could have never walked so far without you by my side. You mean all to me.

The data for this PhD thesis stems from a project that was funded by the Swiss National Science Foundation and supported by the Austrian, German, and Swiss Associations of Occupational Therapy and by an internal grant of the School of Health Professions at the ZHAW.
REFERENCES


References


Bundesversammlung der Schweizerischen Eidgenossenschaft. (2006). *Bundesgesetz für die universitären Medizinalberufe (Medizinalberufegesetz MedBG) [Federal law on the health professions educated at universities]*. Bern: Bundesamt für Bildung und Technik,


References


Der Schweizerische Bundesrat. (2004). *Verordnung über Geburtsgebrechen (GgV) [Regulation on birth defects]*. Bern: Der Schweizerische Bundesrat.


References


Ministrstvo za Zdravstvo. (2005). *Pravila obveznega zdravstvenega zavarovanja [Regulation of the compulsory health insurance]*. Ljubljana: Zalozba GV.
References


Raunig, F. (2010). Mitteleuropa und Identität: Zur Relevanz philosophischen Fragens [Middle Europe and identity: Relevance of a philosophical discourse]. In M. Flatscher, C. Hornstein Tomić, A. Periša & H. Riefenthaler (Eds.), *Zur Relevanz mitteleuropäischer Identität [Relevance of a Middle European identity]* (pp. 7-10). Frankfurt am Main: Peter Langer.


SPSS. (1989-2011). *IBM® SPSS® Statistics* [Computer software] [20.0]: IBM Company.


References


84


