Education and Political Tolerance

A Meta-Analysis of the ”Education Effect”

Author: Johan Peck
Supervisor: Sven Oskarsson

Master’s Thesis, Spring 2016
Department of Government
Uppsala University
# Contents

Introduction .................................................................................................................. 3
  Preamble ...................................................................................................................... 3
  Structure of paper ........................................................................................................ 4
I. Education and political tolerance: An open question? ............................................... 6
  The logic of the education effect .................................................................................. 6
  Previous findings .......................................................................................................... 6
    The Stouffer hypothesis ............................................................................................ 6
    The Sullivan and Jackman critiques .......................................................................... 7
  Divergent data ............................................................................................................. 8
  A research synthesis: Aiming at reconciliation .......................................................... 8
II. Theory and methods: Meta-analysis, key concepts, and operationalisations ............. 10
  Synthesis as a research strategy ................................................................................. 10
    Systematic review ..................................................................................................... 10
  Meta-analysis: The aggregation of quantitative data .................................................. 11
  A research synthesis of the education effect ............................................................... 12
    The effect size ........................................................................................................... 12
  Study characteristics and axes of analysis ................................................................. 13
    Operationalisation of political tolerance ................................................................. 14
  Context of survey ........................................................................................................ 15
  Model specification .................................................................................................... 16
III. Material and data-collection ................................................................................ 17
  Research synthesis data-collection ........................................................................... 17
    Population, sample, and units in research synthesis ................................................ 17
    Reviewing research: Drawing a sample ................................................................... 18
    Criteria of inclusion .................................................................................................. 19
  Material retrieval ........................................................................................................ 19
    Search strategy and delimitations of the search ....................................................... 19
  Database search .......................................................................................................... 20
  Reference searching .................................................................................................... 20
  Criteria of inclusion and exclusion ............................................................................ 20
  Results ......................................................................................................................... 21
  Coding the data .......................................................................................................... 22
  Effect sizes and unique samples ............................................................................... 22
  Tolerance measures .................................................................................................... 23
  Setting of study ........................................................................................................... 23
IV. The aggregated education effect ................................................................. 25
  Fixed and random effects ........................................................................... 25
  Weighting the data ....................................................................................... 26
  Publication bias ............................................................................................. 27
  Aggregated effect sizes ................................................................................. 28
  Size of aggregated effect .............................................................................. 29
  Influence of weighting .................................................................................. 29
V. Analysis of heterogeneity ........................................................................... 31
  Hypotheses for sources of heterogeneity ...................................................... 31
  Pairwise comparisons of mean effect size .................................................... 32
  Meta-regression ............................................................................................ 33
    Meta-regression model specification ......................................................... 34
    Meta-regression results ............................................................................ 35
VI. Discussion .................................................................................................. 36
  Limitations of this study ............................................................................. 36
  The aggregated education effect .................................................................. 36
  Determinants of effect size ......................................................................... 37
References ....................................................................................................... 38
Appendix A: Included samples ........................................................................ 42
Appendix B: Notes on coding and data manipulation ...................................... 43

List of tables
Table 1. Results of search and screening.......................................................... 22
Table 2. Mean effect size ................................................................................ 29
Table 3. Hypothesised influence of study characteristics .................................. 31
Table 4. Pairwise comparisons of means, by study characteristic ....................... 32
Table 5. Weighted multiple regression analysis of effect size estimates .............. 34
Table 6. Included samples .............................................................................. 42
Table 7. Notes on coding and data manipulation .......................................... 43

List of figures
Figure 1. Search process and exclusion criteria .............................................. 21
Figure 2. Funnel plot ....................................................................................... 28
Figure 3. Effect sizes and weighting ............................................................... 30
Introduction

Preamble

Education: The go-to panacea for all social ills. “More education!” has been espoused as a general solution to a great variety of problems which plague the world, and education has been shown to have a positive relationship with most of those phenomena and values which any modern society strives for, from good health (Hahn & Truman, 2015) to general happiness and wellbeing (Michalos, 2008).

Not least of these values are the political and social conditions which are said to promote a vibrant democracy. An educated population makes better, more democratic citizens, with greater respect for democratic ideals, greater political efficacy, and more tolerant attitudes, helping bridge the divides between interests and groups. The last aspect of this broad assumption – that education promotes tolerance of differences, and thus by extension the viability of the democratic model as a system for solving socio-political conflicts and disputes – is the subject of this paper.

In conceptualising tolerance within political science, sociology, social psychology, and related fields, a distinction is often made between social tolerance and political tolerance (e.g. Lee, 2013). The former refers to acceptance in everyday life or society at large of those who somehow stand out from the norm. Be it by different origins, traditions, or the like, or by their lifestyles and values somehow being perceived as marginal. The latter refers to the more specific willingness to extend to these fringe, marginal or disliked groups the political and civil rights generally granted to citizens of a democratic polity. As such, political tolerance is often conceived as a necessary prerequisite for a functioning liberal democracy (Marquart-Pyatt & Paxton, 2007). Only in a society where we allow all citizens equal rights can democracy be vital.

Starting with Samuel Stouffer’s research in the United States in the mid-1950s, political tolerance has therefore grown to a topic of interest unto itself within the field of political science. In this, it has followed other attitude research, usually relying on national or cross-national surveys questioning a population about their views, preferences and opinions. And during this evolution, education has been a constant facet of proposed explanations for varying degrees of political tolerance between and within populations. Stouffer observed that those with more education also held greater levels of political tolerance. This result has later
been present in a large body of survey research, and the relationship has regularly been considered theoretically viable and empirically robust (Van Doorn, 2014: 914-915).

But the suggestion that greater education leads to greater political tolerance has not gone without scrutiny. In fact, critics such as Mary Jackman have, since at least the 1970s, tried to problematise the relationship and show that it is a false correlation, based on the approaches and assumptions of political tolerance research (Jackman & Muha, 1984).

Meanwhile, the observed relationship has at times varied greatly between surveys and statistical models examining the sources of political tolerance. Some researchers have found, sometimes to their surprise, that the education variable of their proposed model is statistically insignificant, substantially null, or even significantly negative (e.g. a study by Sten Widmalm and Sven Oskarsson (2013) found no significant relationship, which surprised the authors).

Therefore, this study seeks more specifically to elucidate the varying relationship between education and political tolerance observable in previous research. It tries to summarise the findings of previous research, aiming towards the hypothetical consensus reached for by most scientific enterprises. At the same time, it tries to explore the variations in, and determinants of, the size and direction of the “education effect”, the quantitative embodiment of the relationship. It does this using the methods of research synthesis and meta-analysis – quantitatively aggregating previous results and re-examining them along with the characteristics of each included study which could affect its results.

While the above stated aims are of course more ideal than realistically achievable, the present research synthesis can hopefully shed some light on the problem at hand. In this, it might offer both some insight into, and critique of, the oft-assumed positive relationship of education on political tolerance. These results may then also apply to assumptions about at least one of the grounds – promotion of tolerance – on which education is conventionally perceived as a normatively attractive value in democratic society.

Structure of paper

The paper is laid out according to the following structure:

It begins with a brief review of the topic at hand: The assumed relationship between education and political tolerance, as well as the key factors motivating a further investigation of this relationship. This concludes in a two-pronged research question.

Then follows a methodological overview of the research strategy employed, the uses of meta-analysis, key concepts, the main axes of inquiry, and their relevance.
Thirdly, a section on the collection of data-material and the coding process.

After that, the data are presented, complete with relevant manipulations and weights, and an analysis of the nature and spread of effects.

Section 5 goes further into the analysis of the education-political tolerance effect, by presenting statistical findings of aggregate study-level relationships, and the key determinants of the size of the effect.

A concluding discussion presents the results in plain text, speculates on implications and further considerations, and anticipates questions for future research.

A note on the description of methods: As methods of research synthesis and meta-analysis may not be familiar to the reader, I have tried to explain each step of the study methodologically, in as plain language as possible. Since this makes the methods-related text quite extensive, I have spread out some of the methodological sections so that they follow the logic and structure of the paper, rather than giving a complete methodological account in one early section. I hope this will facilitate reading. A more general methodological overview is however (as stated above) given in section 2, so the reader may hopefully orient themselves and anticipate the analyses to come.
I. Education and political tolerance: An open question?

The logic of the education effect

The assumption that level of education should influence political tolerance is intuitive to most social scientists on two levels:

Firstly, as referenced above, education is one of the main socio-economic background variables of social science. No matter our topic of research, if we are studying a social phenomenon, we usually think it may be in some way related to or determined by level of education. From this perspective, it is only natural to factor in education when measuring and explaining political tolerance, and to hypothesise on its effect.

Secondly, a positive relationship could easily be expected, based on experience or preconceptions. Intolerance is often espoused as being caused by ignorance, and if education alleviates ignorance, it should also promote tolerance. Educated people “...learn about diverse points of view and ways of experiencing life, thus becoming less dogmatic, even if they had a tendency toward closed-mindedness before achieving a high level of education” (Sullivan & al 1981: 100). In other words, the basic mechanism and theory behind why education should increase political tolerance is for most both easy to understand and to accept at face value.

Previous findings

The Stouffer hypothesis

In a 1955 study often cited as the genesis of political tolerance research, Samuel Stouffer surveyed the American people, asking questions about their willingness to extend civil liberties to people belonging to groups which could be perceived as outside of, or threatening to, mainstream American society (Gibson, 2006: 21).

Apart from other background variables such as location of residence and urbanicity, he found that education had a significant and substantial positive effect on political tolerance. These findings were explained through the mechanics detailed above: Education led to greater political tolerance, he speculated, because those more educated had been exposed to more diversity, and both mentally and practically learned to understand and respect views and people outside their own comfort zone (Bobo & Licari, 1989: 287).

This set the tone for both the expected and observed relationship between education and political tolerance in research to come.
The Sullivan and Jackman critiques

However, the data Stouffer relied on measured political tolerance by asking respondents about their willingness to extend civil liberties to a specific set of groups. And following studies relying on the recurring General Social Survey conducted in the United States have followed this operationalisation. These groups were originally communists, socialists, and homosexuals. This approach was criticised by among others John Sullivan. Sullivan argued that tolerance of these groups was a biased measure: The groups were on the left-wing or progressive side of American politics, and those with higher education were known to have more liberal or left-leaning views. It was therefore not greater tolerance which Stouffer had observed, but merely the differing sympathies of differing educational strata (Sullivan & al 1981).

Sullivan then devised an alternative operationalisation, known as the “least-liked” method. Respondents were asked to select groups they themselves considered objectionable, and then queried on their level of tolerance of these self-selected groups. He then saw the education effect disappear (e.g. Shamir & Sullivan, 1983). The relationship could be spurious, confounded by the sympathies of those interviewed (Sullivan & al 1981: 92). Although the General Social Survey has since then broadened the questions somewhat, it is unclear if its fixed-group approach remains a biased measure.

Meanwhile, Mary Jackman and her fellow researchers questioned the very notion of measuring political tolerance by asking about abstract civil rights. Support for these rights were not necessarily representative of actual tolerant attitudes, but rather a form of socialised façades, more recurrent in higher socio-economic strata and therefore receiving stronger support among those with higher education. And indeed, differing educational effects have been observed when trying to move beyond this, by using for example support for specific “tolerant” policies as a measure of political tolerance (Jackman & Muha, 1984; Vatter, Stadelmann-Steffen & Danaci, 2014).

The education effect then seems to be contingent on the conceptualisation and operationalisation of political tolerance. Among others, James Gibson (1992; 2013) has tried to take these differing operationalisations into account. He has found sometimes conflicting results as to the reliability and consistency of the political tolerance construct when measured in different ways.
Divergent data

A growing body of research, conducted on a greater variety of populations, also found differing effects. When moving outside the U.S. context, results have been far from unequivocal, finding both null and negative relationships (Widmalm & Oskarsson, 2013; Duch & Gibson, 1992). In addition, the nature of political tolerance may be subject to variations over time, as Stouffer’s survey was conducted in not only a very particular place (the U.S.) but also a particular time. That education promoted political tolerance in the mid-20th century does not necessarily mean it does so today.

In addition, studies including education only as a control variable when exploring other determinants, or developing more nuanced and sophisticated models of political tolerance, have also seen the effect vary. These include path-analysis and multiple regression models including previously unaccounted-for variables (e.g. Eisenstein, 2006).

This raises the question if the education effect is dependent not only on the measure of political tolerance, but also contextual factors like time and place of survey, as well as parameters of model specification.

A research synthesis: Aiming at reconciliation

Scientific research aims at iterative accumulation of knowledge, but in this case such accumulation has been stalled by the seemingly paradoxical nature of the relationship between education and political tolerance. From an academical perspective, it seems prudent to untangle the education effect, and paint a both more complete and nuanced picture. This is especially true due to the perfunctory use of education as an independent variable when modelling tolerance, and the uncritical theoretical assumptions behind such use.

At the same time, education is a variable of significant interest in political science in general due to its relative manipulability in concrete policy. From a practical viewpoint, understanding the effects (or non-effects) of education on political tolerance (and its continuation, a democratic citizenry) thus appears highly desirable.

If the variations in the education effect can be accounted for and in turn analysed, perhaps the seeming paradoxes can be resolved. Such resolution would mean a greater understanding of the specifics and nuances of the role education plays in political tolerance. In view of the above, this paper conducts a systematic review of previous findings, in the form of a research review and meta-analysis, to shed light on two basic questions:
Firstly, is there a general tendency in the education effect? How widely does it vary, and is there an aggregate effect with any particular direction? Is that aggregate effect substantially and significantly different from zero?

Secondly, are there any specific study characteristics – operationalisations, context of survey, model specification – which influence the observed relationship? Are there clear directions for such influences, i.e. do any particular study characteristics alter the observed education effect in measurable and statistically significant ways?
II. Theory and methods: Meta-analysis, key concepts, and operationalisations

To answer the questions posed, this study employs, as stated, the tools of research synthesis and meta-analysis. Concretely, this means the collection and review of a body of already conducted primary studies. In this case, studies somehow viewing political tolerance as an outcome, and education as an independent variable. As this approach is perhaps not common currency in political science, a short exposé of this method may first be necessary, with a focus on meta-analysis.

Synthesis as a research strategy

That science is, in its essence, an iterative project is self-evident. By continually building on what we already know, and asking new questions raised by previous research, it seeks to refine and improve our understanding step by step. Yet, more often are new studies conducted in order to examine a topic of interest than effort is invested to actually determine what has been previously accomplished. A research synthesis does the opposite: It systematically reviews previous research, treating primary studies as objects of analysis in themselves.

Systematic review

The phrases “systematic review” and “research synthesis” are roughly equivalent, and are used interchangeably throughout this paper. What differentiates such a review from the normal overview of previous research present in all academical writing is the emphasis on the word “systematic”. While any study reviews previous research when selecting, motivating and specifying its research question and strategy, this is usually done in an arbitrary and ad hoc manner. A researcher simply sifts through their own knowledge, reads and references research at hand which seems relevant, and paints a broad picture of the state of research and the relevance of their own addition.

A systematic review formalises this process. By employing more stringent search strategies, it seeks to collect not only conveniently available or obviously relevant research, but a more or less complete, or at least representative, sample of what has been said on a topic (Cooper & al, 2009: 4-7).
This is then scrutinised and analysed according to some defined standards. Findings are summarised and differences of outcomes problematised. The classical research review does this in a qualitative and narrative way. It relies on the knowledge of the reviewer to find both relevant angles of investigation, and to weigh studies against each other (Cooper & al, 2009: 4-5).

**Meta-analysis: The aggregation of quantitative data**

Meta-analysis is a subgenre of the more general systematic review. Instead of simply gathering and analysing previous research according to some more or less specified strategy, it seeks to use statistical methods to formally aggregate data. A meta-analysis gathers the quantitative output of a body of research, and uses that output as data-points in a new analysis: A statistical study of previous quantitative research (Wolf, 1986: 10-14).

It has seen its greatest use in medical science, as a way to aggregate controlled medical trials. Because of this provenance, common meta-analytical tools often come with a stringent and rigorous set of conditions (Cooper & al, 2009: 10, 88). This has perhaps deterred its use in the social sciences, as these can seldom be met when reviewing e.g. political science research. However, just as political science must employ its own set of tools in primary research, making the best of sometimes unfavourably conditions, the basic logic of meta-analysis can be applied without necessarily using all the pre-packaged gears and cogs of meta-analytical methodology.

In comparison to a qualitative review, meta-analysis has several benefits. Firstly, it does not rely on the knowledge and judgement of an individual reviewer to draw its conclusions. This means it has the same advantages of reliability and reproducibility as quantitative research in general. Secondly, meta-analysis uses a larger amount of data from each reviewed study. A systematic review may, for example, simply count the number of studies which found a significant relationship between two variables, and contrast these with the studies which did not find a significant relationship. This takes into account neither the size of an effect nor the parameters, such as sample size, which could affect its significance. By extracting such data from each study, and combining measurements into a single model, meta-analysis formally tests the size and strength of a relationship (Ellis, 2010: 91-97).

It also allows for the formal inclusion of study characteristics in the analysis. Differences between studies such as time of publishing, methods used, and geographical setting can be coded as study-level variables. These can then be tested together with the main
relationship, to see if an observed effect is contingent on how, when and where it is observed (Cooper & al, 2009: 154-157).

This makes meta-analysis a suitable tool for this study, since it is exactly these questions we seek to answer.

**A research synthesis of the education effect**

Having established the above, how more exactly can meta-analysis explicate the nature of the relationship between education and political tolerance? This can be thought of as a three-step process: First, the gathering of a representative sample of research, and the organising of this sample as a new data-material. Second, the extraction of “effect sizes” (being measurements of the education effect) from each study reviewed, as well as the coding of other study characteristics. (These first two steps are further detailed in section 3.) Thirdly, the aggregation and analysis of effect sizes, and how they are affected by the other variables of interest – the characteristics of each study.

However, before moving on to the choice and operationalisation of study characteristics, the effect size employed is defined.

**The effect size**

The main aggregate statistic of any meta-analysis is the effect size measure. It is therefore worthwhile to dwell somewhat on its conceptualisation and specification.

In its most simple form and in the context of this study, the effect size is the size and direction of the relationship between education and political tolerance. For each study, one or several effect sizes are extracted, representing the education-political tolerance relationship in a given survey sample. These effect sizes may be thought of as a standardized measure of the effect on political tolerance of a given increase in education. Correspondingly, if each survey sample is a unit in our analysis, each effect size is an observation of the variable of interest. Observations which we can then aggregate and analyse with statistical tools.

As regression analysis is the bread and butter of quantitative political science research, it follows that the measure of the effect size in this case must be extracted from such regression models. The effect size is thus merely the regression coefficient of an independent education variable, where the dependent variable is a measure of political tolerance.

However, to ensure that these measures are comparable across studies they must be on the same scale. Otherwise, aggregating the effect sizes – the regression coefficients – would
be impossible. Therefore, the standardized β-coefficient of the education variable in a given model is the measure of preference. This is universally translatable between studies, as it represents the increase in standard deviations of the dependent variable (political tolerance) of each standard deviation increase of the independent variable (education). In other words, it does not matter which scale education or political tolerance is originally measured by: The β-coefficient is in itself standardized to a common metric.

It should be noted that this is not quite as straightforward as stated above. The β-coefficient suffers from some drawbacks, which has limited its use as an effect size measure in previous meta-analyses, and cast some doubt on its suitability. The main problem comes from the fact that regression analysis almost always means multiple regression analysis. And in multiple regression analysis, β-coefficients of independent variables are prone to vary across differently specified models (Becker & Wu, 2007). The interpretation of the meta-analysis may therefore, according to some, be rendered prohibitively difficult (Peterson & Brown, 2005: 175). However, robust variables are affected to a lesser degree by the exact model specification, and education may be considered such a variable in this case.

The aggregation of β-coefficients is also far from unheard of, and recent research has shown that the measure can often be effectively used in meta-analysis – even substituting for the more commonly used Pearson-correlation (Petersen & Brown, 2005). It has also been employed in several research syntheses, ranging from political science (Huang & al, 2009) to medical epidemiology (Nieminen & al, 2013). And apart from this, it also has the added virtue of allowing the model specification – the control variables in each study – to act as another venue of investigation, by coding for the inclusion or exclusion of just such control variables in a study.

Finally, a point made above bears repeating: Meta-analysis in political science may not always be able to – and perhaps should not strive to – live up to the precedence of e.g. methodologically anxious scholars who conduct meta-analysis in the setting of randomised medical trials. While such scholars may not always consider multiple regression coefficients a suitable source of data, this does not invalidate it in the context of quantitative political science.

**Study characteristics and axes of analysis**

Having defined the measure of the education effect, the next step is defining the different study characteristics: The study-level variables which allow for a further analysis of
the variation of the education effect. Following the overview in section 1, three main axes of investigation present themselves: Operationalisations of political tolerance, context of survey, and model specification.

**Operationalisation of political tolerance**

There apparently exists several measures of political tolerance, and the way it is measured seems to influence the effect of education. The first axis along which the education effect may vary between studies therefore seems contingent on the operationalisation of political tolerance.

Gibson (2013) identifies three conceptually distinct measures:

First, Stouffer’s original “fixed-group” tolerance. This measure assumes that a researcher, when constructing a survey, selects outgroups assumed to be objectionable to a population in general. Respondents are then queried through one or more civil liberties questions, usually concerning if a member of an outgroup should be allowed to make public speeches, teach in school, have a book espousing their views in a public library, or the like.

This measure can additionally vary along a sub-dimension: The nature of the outgroups selected. Stouffer’s original study used only outgroups on the left/progressive side of the American socio-political spectrum. Recognising the problems and possible bias of this approach, later surveys have broadened the spectrum to a wider variety of outgroups. This has however not been done consistently. As a result, some surveys employing the fixed-group measure use only left/progressive outgroups (e.g. communists), some use only right/conservative outgroups (e.g. fascists), and some use a mixed selection. This sub-variation must of course be noted and accounted for.

Second, Sullivan’s “disliked-group” tolerance. The questions posed to a respondent are similar as in the fixed-group approach, but the outgroup is self-selected by the respondent. This is done either by the respondent being asked to mention a disliked group, or by being provided with an extensive list of potential groups and being asked to select one or more they particularly dislike. The respondent is then queried about their willingness to extend civil liberties to this disliked group.

Finally, Gibson defines a third measure: General support for civil liberties. This is not measured as political tolerance of any specific group, but rather by asking respondents if all citizens, no matter who they are or what their views are, should be granted equal rights. Since this is not a measure of political *tolerance* in a strict sense, studies employing this method will not actively be searched for in this research review.
In addition, the operationalisation used by e.g. Mary Jackman and Michael Muha (1984) could be considered another relevant measure. This approach queries respondents not about civil liberties in the abstract, but about support for actual legislation which extends the rights of one or more outgroups. It thus tries to capture the gap which may exist between intangible support for equal rights and a concrete willingness to “walk the walk” of tolerance.

However, this measure is rarely captured in indices comparable to those of the more conventional political tolerance research. It is therefore not possible to extract an effect size from such studies. This means the alternative “actual-legislation” measure of tolerance unfortunately cannot be included in this synthesis.

In summary, studies are recorded as employing either a fixed-group or disliked-group measure of political tolerance. In addition, if a fixed-group measure is employed, the outgroups are identified as either only left/progressive, only right/conservative, or mixed.

**Context of survey**

The second axis of analysis is constituted by the context of a given survey. As there is reason to believe the education effect varies across both time and geographical setting, the sampling year and country is recorded for each correspondent effect size.

The temporal aspect is simply the age of an effect, arrived at by subtracting the sampling year from the year of the data-collection (2015), creating a variable for the age of each effect size.

For geographical setting, previous investigation has shown that the relationship between education and political tolerance may be uniquely strong in the United States (Marquart-Pyatt & Paxton, 2007). This is hypothesised relate to the specifics of the college environment in the U.S., and the strong connection between education and citizenry socialisation present in that setting. Therefore, U.S. and non-U.S. studies should show differing effects.

In addition, non-Western countries have shown weaker relationships. This could also be related to the nature of education as an institution socialising prevalent societal norms. In other words, developed liberal democracies would have education systems promoting the norms of political tolerance, while repressive regimes might have the opposite, leading the more highly educated to be less tolerant (Gibson & Duch 1993).

For the purposes of this study, this dimension is captured by the rough approximation of OECD membership. In summary, studies are recorded as being conducted either in the US, a non-US OECD country, or a non-OECD country (at the time of sampling).
Model specification

Finally, the included or excluded control variables in a given model could influence the size of the education effect. Partly, due to the simple mechanics of multiple regression analysis. But also because the education-political tolerance relationship may, as some have suggested, be somewhat or completely spurious.

Therefore, the model specification is the third axis of analysis of the effect size. This is recorded as the additional independent variables, apart from education, present in each model. Specific coding of these study characteristics, and the construction of a data-set suitable for meta-analysis, is further detailed at the end of the next section. First, however, the sampling of the relevant research is described, together with an account of the actual gathering of the material.
III. Material and data-collection

The following section details the logic behind data collection in research synthesis, the search strategy employed in this study, the selection procedure resulting in the final data-material, and the coding of this into a data-set for meta-analysis.

Research synthesis data-collection

Population, sample, and units in research synthesis

In order to explore the general relationship between education and political tolerance, this research review collects and samples previous primary research. In essence, the data-collection methods used in a research synthesis and meta-analysis are analogous to the sampling procedure of normal quantitative research. A survey study, for example, defines a population, and draws an in some way representative sample from that population. It then analyses the sample according to some defined method, and based on the logic and practice of statistical inference, tries to say something about the population at large.

A research synthesis with meta-analysis does the same. But the population in this case is the body of research about a phenomenon, and the sampled units are the included studies which have something to say about that phenomenon. In addition, the total body of research – the population – is defined not only as the studies actually conducted, but a greater universe of potential studies about a subject. And the sample is not, in practice, a random sample of research findings, but usually a more or less complete review of past research (Cooper & al, 2009: 38-44).

In other words, by reviewing research which has actually been done on a subject, a research synthesis and meta-analysis seeks to draw general conclusions about that subject, making inferences about the potential results of all hypothetical research which could be conducted. Given that we were to conduct an additional study on the effect of education on political tolerance, what results could we expect? What are the general results of the reviewed research, valid not only across the included studies but also for the greater universe of research in general, and the examined phenomenon as a whole?

The unit of observation is thus the individual study (or, strictly, a sample with a corresponding observed relationship within each study), the sample is the set of studies selected for review, and the general population are all studies which examine that relationship – real or hypothetical.
Another way to put it is this: In an individual study, the inferences made are strictly speaking valid only for the population sampled. And the relationships observed are contingent on the nature of the study: The time and place it was conducted, the methods used, etc. But what if we want to draw more far-reaching conclusions about those relationships, applicable across time, geographic space, methodological differences, or other delimitations? By widening the horizon and reviewing conducted tolerance research – which varies along one or more of these lines – research synthesis with meta-analysis tries to do just that.

**Reviewing research: Drawing a sample**

Again following the logic of sampling, populations, and statistical inference, the material included in this research synthesis and meta-analysis should be somehow representative of all potential research – corresponding to a sort of random sample. The studies under review must therefore not be biased along lines which could affect the relationship of interest. If only studies with statistically significant results are included, for example, then the relationship will probably be over-estimated (Cooper & al, 2009: 436).

The common method of achieving an unbiased data-material is to try to take into account all results studying a relationship. If this is achieved, the results from a meta-analysis should in turn be valid across all settings – with the (quite important) caveat that there may be types of studies which have not yet been conducted (Cooper & al, 2009: 43). This means including an as complete as possible body of published articles (making sure that there is no “publication bias” – the tendency to mainly publish significant results in journals, further explored in section 4) by searching bibliographical databases, looking up references, etc. It also means finding and integrating so called “grey literature”: Conference papers, books, unpublished research articles, non-academical studies of scientific quality (e.g. government reports) and the like (Cooper & al, 2009: 105-106). All this literature is then reviewed for inclusion or exclusion based on criteria relevant to the relationship being studied and the nature of the research synthesis.

When a complete sampling is not possible, a representative one may be pursued instead. This means delimiting the search along lines which are considered meaningful or relevant. As the sample is delimited, the population about which one can make inferences is correspondingly narrowed down. While actual random sampling is sometimes employed if a body of research is very extensive, the material is usually narrowed down by substantive inclusion criteria and/or practical search delimitations rather than by setting a fixed sample size in advance.
Criteria of inclusion

As for any sample, criteria of inclusion must be defined in a way that is relevant and mirrors the characteristics of the population. As the ultimate goal is to say something about a specific relationship, studies should first and foremost be eligible for inclusion if they examine that relationship in a relevant way.

Commonly, this means including studies which fulfil one or more methodological and results-based criteria. For a meta-analysis, the basic criterion of inclusion is that a study reports an effect size which is comparable to other effect sizes and possible to aggregate. Furthermore, criteria of inclusion may comprise methodological aspects such as only including experimental studies, estimates of study quality, or context-based criteria such as only including studies conducted in a specific setting (Cooper & al, 2009: 161-163).

Material retrieval

For the purposes of this study, a search strategy was developed to result in a substantively relevant and pragmatically manageable material. This strategy comprises both the procedure for searching the political tolerance literature, and the criteria of inclusion.

Search strategy and delimitations of the search

Within the confines of this limited paper, a complete review of both published and grey literature is of course not feasible. Instead, the search strategy employed tries to capture the most relevant results given limited time and resources for reviewing the literature.

Firstly, only published scientific articles were searched for and considered eligible for inclusion. This means that the conclusions later drawn are mainly relevant to that type of research. However, since scientific articles comprise the vast majority of reported results, this delimitation should not greatly limit the ability to draw conclusions about the education-political tolerance relationship itself.

Secondly, the search was limited to a two-stage process: First, an online search of bibliographical databases, and then a second-stage a snowball search, where the references of the initially included articles were reviewed.
Database search

The first stage searched the following databases, being some of the most relevant political science bibliographies, for the subject term “political tolerance”:

- International Bibliography of the Social Sciences
- PAIS International
- Worldwide Political Science Abstracts
- Sociological Abstracts

The search results of these searches were reviewed in full. Pilot searches of other bibliographies such as the Social Science Citation Index did not return significantly different results. The pilot searches’ results in every case mirrored the previous searches to a very large degree, and a partial screening did not result in any new articles eligible for inclusion. Due to pragmatic limitations, these further search results were therefore not systematically reviewed.

Reference searching

After the first-stage results had been screened (this screening process is explained below), the articles included had their references reviewed for further items of relevance. Referenced articles were examined in full if the original referent either: (a) seemed reliant on the reference in question, or (b) if a title in the reference list was deemed promising, e.g. by suggesting the reference was indeed an article on political tolerance. This was not a completely systematic process, but conducted as thoroughly as limitations allowed.

Criteria of inclusion and exclusion

A two-stage screening process was used to judge eligibility of inclusion. For an illustration of the retrieval and exclusion process, see Figure 1.

Items retrieved through the bibliographical search were first screened for any possible relevance. This in practice meant reading the abstracts of articles, and in some cases the full text. The only criteria of inclusion was that the article must contain a quantitative estimate of the effect of education on political tolerance. Those which did not contain, or did not seem to contain, such an estimate were excluded.

Items which were not research articles, such as book reviews, scientific books, conference papers etc., were as stated above excluded. Additionally, articles not written in English were excluded due to practical concerns, as were articles which were not accessible online. Finally, research published before 1980 was excluded due to the problems of irretrievability and the differing standards of scientific publishing of earlier research.
The second screening meant reading the articles in full, which also applied to any identified promising references. Articles were then excluded if they fulfilled one or more of the following exclusion criteria:

a) The article did not, upon closer examination, in fact report a quantitative relationship between education and political tolerance.

b) The statistical model did not report a coefficient convertible to, or comparable with, the effect size measure (regression output $\beta$-coefficients).

c) The study relied on the same survey sample as another study. For criteria of selection between such same-sample studies, see Appendix B: Notes on coding and data manipulation.

The articles not excluded due to any of the above reasons constitute the final sample.

**Results**

The results of the search are shown in Table 1 below. In total, 141 items were identified through either bibliographical search or reference searching. Of these, 56 passed first-stage screening and were reviewed in full in second-stage screening.
Table 1 also presents the number of excluded articles in the second screening. After more thoroughly studying the articles, 20 remained which contained data eligible for meta-analysis. These 20 articles however do not constitute just 20 data-points. Several articles contain multiple distinct samples and effect sizes, see Coding the data, below. A complete list of units of analysis is presented in Appendix A: Included samples.

**Table 1. Results of search and screening**

<table>
<thead>
<tr>
<th>Bibliographical search</th>
<th>Reference search</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items reviewed</td>
<td>1&lt;sup&gt;st&lt;/sup&gt; screen positives</td>
</tr>
<tr>
<td>107</td>
<td>22</td>
</tr>
<tr>
<td>Total first screening positives:</td>
<td>56</td>
</tr>
</tbody>
</table>

**Second screening**

Excluded due to:

- No quantitative relationship: 18
- No usable effect size: 17
- Duplicate sample: 1

Total second screening positives: 20

**Coding the data**

The following section outlines the construction of the data-set. A more complete description of coding and data manipulation can be found in Appendix B: Notes on coding and data manipulation.

**Effect sizes and unique samples**

As is common in reporting of quantitative research, most studies included more than one model describing their findings. However, a unique effect size cannot be derived from each of these, as they are in many cases based on the same survey data. Therefore, it is the number of unique survey samples reported within a study which determines the number of observations it provides to the data-set. For criteria of which model and effect size was recorded when several relied on the same sample, see Appendix B: Notes on coding and data manipulation.
For each unique sample and model, ranging from 1 to 12 per article, the main effect size was extracted from that models corresponding \( \beta \)-coefficient of the education variable, or any measurement convertible or comparable to this measure. In addition, the sample size of each sample was retrieved. This resulted in 43 unique units of analysis.

From here on, these samples and their effect sizes are considered the units of analysis – “sample” and “unit” is in the following text used somewhat interchangeably to refer to the units of the data-set.

**Tolerance measures**

The tolerance measure was coded as either “fixed-group” or “disliked-group” tolerance, creating a dummy variable for disliked-group tolerance, with fixed-group tolerance serving as the reference category.

If a study proposed to measure political tolerance, but proved to use the “civil liberties” approach detailed above, this was coded together with fixed-group tolerance.\(^1\)

In addition, samples with fixed-group measures were coded as using either right/conservative, left/progressive or mixed outgroup/s. A dummy variable was created for the right/conservative measure, with mixed and left/progressive serving as the reference category.\(^2\)

**Setting of study**

Age of sample was coded, as stated, coded as the number of years since the survey was conducted, with 2015 being the reference year.

Samples were also coded as either originating in the U.S., a non-U.S. OECD country, or a non-OECD country, creating two dummy variables for the latter two groups, with the U.S. as the reference category.

---

\(^1\) This applied to two samples. It is motivated by the fact that the main theoretical divide in the operationalisation of political tolerance runs between tolerance of outsiders in general, and tolerance of self-selected disliked groups.

\(^2\) The reason for combining the mixed and left/progressive groups is that actual coding revealed only one study used a fixed tolerance measure with only a left/progressive outgroup. To increase statistical power the original dummy variable for left/progressive outgroup was dropped and this observation included in the reference category.
Model specification

Finally, the model specification was recorded. A dummy variable was created for each of the ten most commonly observed independent variables used together with education. These are:

- **Sex**, if the model controls for sex or gender of respondents.
- **Age**, if the model controls for age of respondents.
- **Ideology**, if the model accounts for ideology or party sympathy.
- **Religiosity**, if level of religiosity is controlled for.
- **Income**, representing controls for both actual income and proxies of economic status such as house-ownership.
- **Support for democracy**, if such a metric is included.
- **Personality**, representing controls for personality type, tendency towards authoritarianism, or similar.
- **Urbanicity**, if a measure of the size of respondents’ place of residence is included.
- **Political knowledge**, including both specifically political knowledge but also related constructs such as political efficacy.
- **Threat level**, if the model controls for perceived level of threats from outgroups or in society in general.

This data retrieval and coding resulted in the final data-set, used in the analysis below.
IV. The aggregated education effect

In this section, the data is tested for publication bias and basic results of aggregating the effect sizes are presented. First, however, considerations of model assumptions and weighting of data must be accounted for.³

Fixed and random effects

Two main models are usually employed in meta-analysis for aggregating and analysing effects: The fixed effects and random effects models.

The fixed effect model, as the name implies, supposes a global, fixed effect size. The effect size is assumed to have a “real”, constant value, present in “the world out there”. This may then vary from study to study, being affected by study-level variables (Ellis, 2010: 127-128). But the residual spread of effect sizes – the variability after taking differences between studies into account – is assumed to be due to random measurement error. Mathematically, this can be expressed by the formula (Huang & al, 2009: 457):

$$T_i = \theta + \epsilon_i$$

Where $T_i$ is an observed estimated effect size, $\theta$ is the assumed global effect size, and $\epsilon_i$ is the random error of the estimated effect size. If we take study-level variables into account, as will be done in the section 5, this expands to (Huang & al, 2009: 458):

$$T_i = x_i\Delta + \theta + \epsilon_i$$

Where $x_i$ are the values of study level characteristics and $\Delta$ is the size and direction of their influence on the effect size. In other words, the differences in estimated effect size due to differences between studies, or study heterogeneity.

The random effects model, on the other hand, presumes that effect sizes may inherently vary between studies. It adds an unknown term of variability to the aggregate effect size measure. Each effect size is then considered an observation which measures a unique actual effect, so that the population effect size $\theta$ differs for each sample (Cooper & al, 2009: 296-297). These unique population effects are then assumed to be normally spread around a super-mean, often designated $\mu$.

³ All statistical analysis was carried out in Stata 12, with an added package for meta-analytical functions.
In practice, this adds another level of uncertainty to the meta-analysis, defined by the spread of observed effect sizes $T$ and the within-study variance associated with each observed $T$. The benefits of this, apart from the theoretical motivation for accounting for possible variation in population effects, is that it diminishes the risk of making a Type I error when aggregating effects and e.g. testing the statistical significance of an observed effect (Ellis, 2010: 130).

In the end however, the choice of model comes down to partly theoretical considerations, and partly practical limits. This study will employ a fixed effects model. Theoretically, this is motivated by the fact that the research question as such assumes such a global effect, and that study heterogeneities should account for the differences in observed effect sizes. Put differently, the very aim of this study is not merely to arrive at a (perhaps overly conservative) estimate of the relationship between education and political tolerance. It is to conduct an exploration of the study differences which could influence such an effect, and to explain the differences in observed effect sizes. The fixed effects model is then logically motivated. The downside to this is that the fixed effects model cannot strictly make inferences about unknown studies (Cooper & al, 2009: 547-548).

Practically, the fixed effects model is also the necessary choice due to lack of reporting of study results in the included studies. The standard errors of effect size estimates, necessary when actually computing the random effects model, are not always available.

**Weighting the data**

Not all studies are conducted equally, and not all results are to be considered equally valid or reliable. The logic of giving higher credence to a high-quality study conducted on a large sample, compared to a brief summary of a small-sample probe, is intuitively apparent (Wolf, 1986: 39). It can therefore be tempting to strongly weight data in favour of those studies found to be of higher scientific quality, employing more valid measures or better sampling strategies, or simply basing their results on a larger sample with greater reliability of estimators. Different weights can thus be employed when aggregating and synthesising effect size measures, as well as when conducting various meta-analytical analyses (Cooper & al, 2009: 259-261).

However, weighting data can also lead to unforeseen and sometimes hard-to-understand results and, potentially, biases. This concern is not to be taken lightly, and weighting should be done with some caution, as well as a thorough understanding of the logic and expected
implications of data manipulation. This is not least a problem when weighting for qualitative variables such as the perceived “scientific quality” of a study. Different weighting schemes, as well as an unquantifiable degree of subjectivity in categorising data, may lead to increased rather than reduced bias (Cooper & al, 2009: 260-261). Because of this, the present meta-analysis does not employ any such “quality” weighting, or any similar scheme based on the apparent nature of a given study.

But the logic behind weighting by study size or similar statistical parameters is more intuitive, and the interpretation more clear. If we presume to aggregate data in hopes of finding a common effect size, the sample size of each study in the synthesis is obviously of concern, as a larger study will more accurately estimate such a common effect. This is commonly done by one of two methods: Weighting by ether within-study sample size (n), or within-study effect size variance\(^4\) (Ellis, 2010: 102-103).

These measures are intrinsically linked, as the variance is itself dependent on the sample size. While the second weight is more nuanced, and usually perceived as the weight of choice in meta-analysis, the above mentioned lack of reported standard errors makes it impossible in this case. Sample size weights are therefore used in this study, when applicable. However, when relevant, both weighted and unweighted results are presented.

**Publication bias**

The first test of the data should be a check for publication bias. Such bias is the result of the tendency of researchers and journals to submit and publish significant findings more often than insignificant ones, and substantial results more often than null effects. Publication bias is a well-known and often prevalent problem in research synthesis, and if it is detected, data should be manipulated to reduce its effects (Cooper & al, 2009: 428-431).

The easiest way to detect this is to plot the effect sizes against sample size. This scatter plot variant is known as a funnel plot, because the plotted effect sizes ideally form a full, upside-down funnel shape when no publication bias is present: A kind of pyramid, centred at the mean effect size. The logic behind this is simple. Studies with greater sample size should have a more accurate estimate of the true effect size, forming the top of the funnel, while studies with smaller samples should have a greater spread of estimates, forming the wider base (Egger & Smith, 1998).

---

\(^4\) More specifically, the inverse square of the standard error (SE\(_\beta\)) of the effect size estimator (\(\beta\)), or \(\frac{1}{SE^2}\).
Publication bias most commonly takes the form of small-sample studies with small effects – effects which might be statistically insignificant – being conspicuously missing. This shows as a chunk being taken out of the base of the funnel (Cooper & al, 2009: 428-429).

Figure 2 presents the result of a funnel plot. While the observations do not form a full pyramid, as n-values over 2000 are few, the general spread is symmetrical and roughly funnel-shaped, apart from one large-sample and high-effect observation and one small-sample low-effect observation. There seems to be no tendency to not report small-sample low-effect results, which is the main object of concern.

As such, publication bias does not seem to be present. A post hoc explanation for this is perhaps that the size of education effect has not been a main concern in the included studies. Being multivariate studies of political tolerance, the lack of a significant education coefficient has probably not deterred researchers or journals from publication.

### Aggregated effect sizes

The first step of the meta-analysis proper is a calculation of the mean effect size of all samples. These aggregated means are also tested to see if they are significantly different from
0. That is, if there at all is an observable and statistically significant effect of education on political tolerance.

**Size of aggregated effect**

Table 2 presents the aggregated means of unweighted and sample size-weighted effect sizes and their associated standard errors, SE. The p-value is the level of significance for a test if the mean effect size is different from zero.

The aggregate effect size has a positive value, and is highly statistically significant. There thus seems to be an actual positive relationship between education and political tolerance when summarising previous research.

The mean standardised effect is roughly 0.1. Substantially, this is certainly a non-trivial effect. In most countries, one standard deviation in schooling years is 2.5-3.3 years (Huang & al, 2009: 457). This is comparable to going from completing primary to completing secondary school in most systems, or from secondary school to getting a basic university or college degree. Such an increase in education would then account for an increase in political tolerance by roughly 10 percent of its standard deviation.

**Table 2. Mean effect size**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SE of mean</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unweighted ES</td>
<td>0.082</td>
<td>0.017</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Weighted ES</td>
<td>0.106</td>
<td>0.018</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

**Influence of weighting**

The larger mean effect size of the weighted data shows that some studies with greater sample size also display larger effect sizes. To understand how this weighting affects the estimate, effect sizes are plotted in Figure 2, with size of marker representing relative weight when calculating the mean.

As can be seen, there is mainly one large sample (also identified in the funnel plot above) which adds to the weighted mean being higher than the unweighted. Looking closer at this observation, taken from a study by Allan Cigler and Mark Joslyn (2002), the model from which the effect size is extracted does not stand out on any theoretical or methodological level. The reason for the large sample size is a pooling over several years of different samples, but this age variable is controlled for in the model. There is therefore no theoretically sound reason to exclude the observation as an outlier or prefer an unweighted mean.
### Figure 3.
Effect sizes and weighting. Relative weight given to each study when calculating weighted mean. Size of marker represents weight. Plotted line at weighted mean effect size.

<table>
<thead>
<tr>
<th>Study ID</th>
<th>% Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bobo &amp; Licari 1989</td>
<td>1.92</td>
</tr>
<tr>
<td>Djupe &amp; Calfano 2013</td>
<td>0.49</td>
</tr>
<tr>
<td>Duch &amp; Gibson 1992 (A)</td>
<td>1.10</td>
</tr>
<tr>
<td>Duch &amp; Gibson 1992 (B)</td>
<td>1.04</td>
</tr>
<tr>
<td>Duch &amp; Gibson 1992 (C)</td>
<td>1.43</td>
</tr>
<tr>
<td>Duch &amp; Gibson 1992 (D)</td>
<td>1.27</td>
</tr>
<tr>
<td>Duch &amp; Gibson 1992 (E)</td>
<td>0.69</td>
</tr>
<tr>
<td>Duch &amp; Gibson 1992 (F)</td>
<td>1.24</td>
</tr>
<tr>
<td>Duch &amp; Gibson 1992 (G)</td>
<td>1.07</td>
</tr>
<tr>
<td>Duch &amp; Gibson 1992 (H)</td>
<td>0.37</td>
</tr>
<tr>
<td>Duch &amp; Gibson 1992 (I)</td>
<td>1.22</td>
</tr>
<tr>
<td>Duch &amp; Gibson 1992 (J)</td>
<td>1.07</td>
</tr>
<tr>
<td>Duch &amp; Gibson 1992 (K)</td>
<td>0.82</td>
</tr>
<tr>
<td>Duch &amp; Gibson 1992 (L)</td>
<td>0.91</td>
</tr>
<tr>
<td>Froese, Bader &amp; Smith 2008</td>
<td>1.92</td>
</tr>
<tr>
<td>Golebiowska 2006</td>
<td>3.09</td>
</tr>
<tr>
<td>Gibson 1987</td>
<td>2.78</td>
</tr>
<tr>
<td>Gibson 1992</td>
<td>0.90</td>
</tr>
<tr>
<td>Gibson &amp; Duch 1993 (A)</td>
<td>2.74</td>
</tr>
<tr>
<td>Gibson &amp; Duch 1993 (B)</td>
<td>3.09</td>
</tr>
<tr>
<td>Karpov 1999 (A)</td>
<td>3.01</td>
</tr>
<tr>
<td>Karpov 1999 (B)</td>
<td>3.43</td>
</tr>
<tr>
<td>Iglic 2010 (A)</td>
<td>2.01</td>
</tr>
<tr>
<td>Iglic 2010 (B)</td>
<td>2.31</td>
</tr>
<tr>
<td>Iglic 2010 (C)</td>
<td>1.86</td>
</tr>
<tr>
<td>Iglic 2010 (D)</td>
<td>3.00</td>
</tr>
<tr>
<td>Iglic 2010 (E)</td>
<td>1.45</td>
</tr>
<tr>
<td>Iglic 2010 (F)</td>
<td>3.92</td>
</tr>
<tr>
<td>Iglic 2010 (G)</td>
<td>1.35</td>
</tr>
<tr>
<td>Iglic 2010 (H)</td>
<td>1.55</td>
</tr>
<tr>
<td>Iglic 2010 (I)</td>
<td>1.30</td>
</tr>
<tr>
<td>Iglic 2010 (J)</td>
<td>7.47</td>
</tr>
<tr>
<td>Gibson 2013</td>
<td>1.70</td>
</tr>
<tr>
<td>Peffley, Knigge &amp; Hurwitz 2001</td>
<td>3.23</td>
</tr>
<tr>
<td>Rhodes 2012</td>
<td>1.51</td>
</tr>
<tr>
<td>Lee 2013</td>
<td>1.57</td>
</tr>
<tr>
<td>Tsang 2013</td>
<td>11.23</td>
</tr>
<tr>
<td>Cigler &amp; Joslyn 2002 (A)</td>
<td>2.92</td>
</tr>
<tr>
<td>Cigler &amp; Joslyn 2002 (B)</td>
<td>0.90</td>
</tr>
<tr>
<td>Golebiowska 1995</td>
<td>0.92</td>
</tr>
<tr>
<td>Seligson &amp; Caspi 1983</td>
<td>2.77</td>
</tr>
<tr>
<td>Petersen &amp; al 2011</td>
<td>10.75</td>
</tr>
<tr>
<td>Weldon 2006</td>
<td>100.00</td>
</tr>
<tr>
<td>Overall</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Effect size range: -0.3 to 0.4
V. Analysis of heterogeneity

This section analyses the observed heterogeneity in effect sizes. It tries to account for the differing effects in samples, by looking at the study-level variables identified and coded for each sample.

Hypotheses for sources of heterogeneity

The direction of influence of study characteristics on the association of education and political tolerance is expected to follow certain patterns, formulated here as informal hypotheses.

Firstly, due to changing demographic and socioeconomic structures, it is expected that age of sample is positively correlated with effect size. That is, that the association between education and political tolerance was stronger in the past than it is today.

Secondly, the effect size is expected to be greater in the US than the rest of the world, and greater in OECD countries than non-OECD countries.

Thirdly, it is expected that observations with a disliked group method of measuring tolerance show weaker effects from education, following Sullivan’s and Jackman’s original critiques (see section I). It is also expected that observations with a fixed right-wing outgroup show a lesser effect, since the purported left-wing tolerance bias of the educated should not make them tolerant of right-wing groups.

Table 3. Hypothesised influence of study characteristics

<table>
<thead>
<tr>
<th>Study characteristic</th>
<th>Expected influence on education effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>↑</td>
</tr>
<tr>
<td>Non-US sample</td>
<td>↓</td>
</tr>
<tr>
<td>Non-OECD sample</td>
<td>↓</td>
</tr>
<tr>
<td>Disliked group tolerance measure</td>
<td>↓</td>
</tr>
<tr>
<td>Right-wing outgroup</td>
<td>↓</td>
</tr>
<tr>
<td>Sex control</td>
<td>?</td>
</tr>
<tr>
<td>Age control</td>
<td>?</td>
</tr>
<tr>
<td>Ideology control</td>
<td>↓</td>
</tr>
<tr>
<td>Religiosity control</td>
<td>?</td>
</tr>
<tr>
<td>Income control</td>
<td>↓</td>
</tr>
<tr>
<td>Support for democracy control</td>
<td>↓</td>
</tr>
<tr>
<td>Personality control</td>
<td>↓</td>
</tr>
<tr>
<td>Urbanicity control</td>
<td>↓</td>
</tr>
<tr>
<td>Political knowledge/efficacy control</td>
<td>↓</td>
</tr>
<tr>
<td>Perceived threat level control</td>
<td>↓</td>
</tr>
</tbody>
</table>
Fourthly, it is expected that the inclusion of any or all types of socio-economic or political control variables should lessen the effect of education. All these are expected to have the same direction in their associations with both education and tolerance, thus accounting for a degree of the education-political tolerance relationship if controlled for. E.g. urbanicity should be positively correlated with both education and tolerance, and thus dampen the direct effect of education if controlled for. The possible effect of controlling for sex, age and/or religiosity, however, is unclear.

See Table 3 for a summary of hypothesised influence on the education effect for each study characteristic.

**Pairwise comparisons of mean effect size**

A simple test of the influence of the selected study characteristics is to compare the mean effect sizes of those with and without the particular characteristic, following Huang & al (2009: 458). The difference in means between each paired group is also tested for statistical significance with a simple two-tailed t-test. This is done for each characteristic except age of sample, which is not a dichotomised variable. Table 4 displays the results.

**Table 4. Pairwise comparisons of means, by study characteristic**

<table>
<thead>
<tr>
<th>Study characteristic</th>
<th>Obs.</th>
<th>Mean</th>
<th>Dif. in means</th>
<th>Sig. of dif.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-US sample</td>
<td>31</td>
<td>0.042</td>
<td>0.143</td>
<td>0.00</td>
</tr>
<tr>
<td>Non-OECD sample</td>
<td>9</td>
<td>0.012</td>
<td>0.088</td>
<td>0.02</td>
</tr>
<tr>
<td>Disliked group tolerance measure</td>
<td>7</td>
<td>0.048</td>
<td>0.041</td>
<td>0.20</td>
</tr>
<tr>
<td>Right-wing outgroup</td>
<td>13</td>
<td>0.088</td>
<td>-0.009</td>
<td>0.40</td>
</tr>
<tr>
<td>Sex control</td>
<td>38</td>
<td>0.075</td>
<td>0.062</td>
<td>0.13</td>
</tr>
<tr>
<td>Age control</td>
<td>42</td>
<td>0.083</td>
<td>-0.063</td>
<td>-</td>
</tr>
<tr>
<td>Ideology control</td>
<td>38</td>
<td>0.072</td>
<td>0.088</td>
<td>0.05</td>
</tr>
<tr>
<td>Religiosity control</td>
<td>36</td>
<td>0.081</td>
<td>0.082</td>
<td>0.47</td>
</tr>
<tr>
<td>Income control</td>
<td>29</td>
<td>0.113</td>
<td>-0.096</td>
<td>0.00</td>
</tr>
<tr>
<td>Support for democracy control</td>
<td>18</td>
<td>0.075</td>
<td>0.012</td>
<td>0.37</td>
</tr>
<tr>
<td>Personality control</td>
<td>9</td>
<td>0.079</td>
<td>0.004</td>
<td>0.47</td>
</tr>
<tr>
<td>Urbanicity control</td>
<td>13</td>
<td>0.016</td>
<td>0.094</td>
<td>0.01</td>
</tr>
<tr>
<td>Political knowledge/efficacy control</td>
<td>29</td>
<td>0.038</td>
<td>0.136</td>
<td>0.00</td>
</tr>
<tr>
<td>Perceived threat level control</td>
<td>20</td>
<td>0.071</td>
<td>0.020</td>
<td>0.29</td>
</tr>
</tbody>
</table>

“Obs.” is number of observations with the characteristic. “Mean” is effect size mean of these. “Dif. in means” is the difference in means between these and all other observations. “Sig. of dif.” is the significance of difference in means. Unweighted means.
It should be noted that when conducting a large number of tests for significance, the risk of a type I error increases if the threshold for significance is not adjusted accordingly.

All characteristics were expected to deflate the association between education and political tolerance, but three instead show a larger effect size (negative difference in means). The influence of using a fixed, right-wing outgroup, however, is both substantially trivial and statistically insignificant, and the influence of including a control for age is meaningless since only one observation does not include such a control.

It is however interesting to note that controlling for income has a substantial and statistically significant positive effect on the effect size (negative dif. in means). This means that studies which control for income will show a stronger relationship between education and political tolerance. Since income is almost certainly positively correlated with education, that means it is negatively correlated with political tolerance – a perhaps unexpected result.

Among the other study characteristics, the results are more expected. Moving further from a US setting decreases the education effect, substantially and statistically significantly. Using a disliked group measure of political tolerance also has a substantial impact, though the statistical significance is weak in this test.

Including a control for sex also deflates the education effect, but again not significantly. It is also unclear why this should be, and the substantial difference may be a statistical artefact. Controlling for religiosity, personality, support for democracy or perceived threat has no significant impact. These indicators, related to outlook, lifestyle and psychology, seem of little consequence.

Socioeconomic and –political controls for ideology, urbanicity and political efficacy, however, have a stronger impact. The weakest among these is ideology, which is also the most personality-related of the variables, and it borders on being statistically insignificant.

In total, the geographical setting of a study, and whether it accounts for socioeconomic factors which may be related to both education and political tolerance, seem to influence the education effect the most when only controlling for one study characteristic at a time.

The next step is to construct a theoretically sound model including several variables, to more completely account for observed heterogeneity.

**Meta-regression**

Much like multiple regression is the statistical tool of choice in primary research, meta-analysis can employ meta-regression to analyse the determinants or influences on the effect
size. This requires weighting the data (in this case using the sample sizes) and, as always when conducting regression analysis, developing a viable model.

**Meta-regression model specification**

The quite low number of observations in this study makes regression analysis a somewhat limited tool. Therefore, including all variables in a large model will not yield interpretable or constructive results. Including many variables in a model with few observations will lead to only variables which contribute an extraordinary amount of explanatory power (by a large $r^2$) to show statistical significance (Milton, 1986: 113-115).

A first model is therefore specified, based on the theoretically most viable variables for explaining differences in effect size (and, admittedly, in part on the findings of the previous tests above). These are the setting, where the important variable is if the sample is a non-US sample; the methodology, best captured by the disliked outgroup-variable; socioeconomic factors, captured by the presence or absence of an income control; and socio-political factors, captured by the presence or absence of a political knowledge/efficacy control.

In addition to these, a larger model is tried which also includes the age of the sample; the presence or absence of an urbanicity control; and whether there is a personality/psychology control in the form of the most theoretically relevant variables: ideology control and perceived threat control. All these have theoretical merit and could well affect an observed education-political tolerance relationship.

**Table 5. Weighted multiple regression analysis of effect size estimates**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-US sample</td>
<td>-.114 (.025)**</td>
<td>-.121 (.038)**</td>
</tr>
<tr>
<td>Disliked group tolerance measure</td>
<td>-.100 (.024)**</td>
<td>-.085 (.031)**</td>
</tr>
<tr>
<td>Income control</td>
<td>.069 (.026)*</td>
<td>.027 (.039)</td>
</tr>
<tr>
<td>Political knowledge/efficacy control</td>
<td>-.070 (.027)*</td>
<td>-.067 (.030)*</td>
</tr>
<tr>
<td>Age of sample</td>
<td></td>
<td>.002 (.001)</td>
</tr>
<tr>
<td>Ideology control</td>
<td>-.029 (.041)</td>
<td></td>
</tr>
<tr>
<td>Urbanicity control</td>
<td>-.021 (.034)</td>
<td></td>
</tr>
<tr>
<td>Perceived threat level control</td>
<td>.017 (.039)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>.184 (.029)**</td>
<td>.206 (.059)**</td>
</tr>
<tr>
<td>Adjusted R$^2$</td>
<td>0.70</td>
<td>0.69</td>
</tr>
<tr>
<td>n</td>
<td>43</td>
<td>43</td>
</tr>
</tbody>
</table>

*Unstandardized regression coefficients, standard errors within parenthesis. Observations weighted by sample size. *p<0.05 **p<0.01 (two-tailed test)*
Meta-regression results

The second, expanded model adds no explanatory power in accounting for the heterogeneity of effect sizes, compared to the first, smaller model (adjusted $R^2$ is in fact slightly lower). None of the four added variables display substantially noteworthy or statistically significant coefficients. The expanded model is thus not any improvement on the smaller model.

The smaller model seems well specified and accounts for a relatively large amount of observed variation. Adjusted $R^2$ is 0.70, which may actually be considered an indicator of a very well fitted model, given the circumstances. All included variables are significant at the $p<0.05$ level, and the Non-US sample variable as well as the Disliked outgroup variable are significant at the $p<0.01$ level. The coefficients are also non-trivial – all else equal, a non-US study is expected to display an effect size 0.114 lower than a US study, a certainly substantial decrease in the estimated education effect, given that the estimated effect sizes vary from roughly -0.3 to 0.3. The Disliked outgroup coefficient is in the same size range. All else equal, simply deciding on a self-selected rather than pre-selected outgroup operationalisation of political tolerance is expected to greatly affect the results of a study.

The fact that the presence or absence of an income control has a substantial and statistically significant effect even in the multivariate model is again interesting as this gives further credence to the proposition that economic status (which the income control variable represents) is negatively correlated with political tolerance. However, the mechanisms behind this finding must be traced in primary research, and one probably shouldn’t draw far-reaching conclusions.

That the inclusion of a political knowledge/efficacy control would deflate the education effect is intuitively understandable, and a statement that seems to hold up in the multivariate analysis. Education and political knowledge are probably both strongly correlated and somewhat over-lapping constructs.

It is also possible to interpret the constant. This is the expected effect size or a study which is conducted in the US, using a fixed outgroup political tolerance measure and no controls for economic status or political knowledge or efficacy. The point estimate of the effect size for such a study configuration is 0.184, in the upper range of observed effect sizes.
VI. Discussion

This study has tried to conduct a limited research synthesis and meta-analysis of the relationship between education and political tolerance. Such a consolidation of previous findings is motivated by the differences in observed effects in the literature. These differences have implications both for researchers in the field, and what considerations must be made when conducting political tolerance research, as well as for the practical real-world applications and interpretations of research results. The results should to a certain extent speak for themselves, but here follows a short discussion of observations.

Limitations of this study

Some limitations of this study must be noted. Firstly, primary research in the field proved to be both so diverse and results so scantily reported that only a small number of studies were possible to extract data from. This is of course a problem when trying to draw conclusions about the field as a whole.

Secondly, again due to limited reporting of results, data manipulations and less than ideal instruments of analysis had to be employed. A rigorous review of the statistical tools would probably find fault with this synthesis, at least if compared to the standards of meta-analysis in medical science or psychology.

Thirdly, the actual collection of data was somewhat limited by the small size of the study and my own limited knowledge of the field. For example, grey literature was not searched at all.

Still, hopefully these limitations have not completely invalidated the results. As stated previously, political science cannot always apply the rigorous methods of other fields. And a limited meta-analysis is perhaps at least a step in the right direction, for both our understanding of political tolerance, and the uses of meta-analysis in political science.

The aggregated education effect

It seems there is indeed a positive relationship between education and political tolerance. This relationship however, is not as strong as some previous research has suggested. It is clear that it is also to a large degree conditional upon study characteristics.
Researchers should thus not simply assume a positive relationship. The same goes for those who would suggest that more education will automatically make people more tolerant. Such a general claim seems very weak, in light of this study.

**Determinants of effect size**

The main aim of this study was to explore how study characteristics may affect the observed effect of education on political tolerance. It turns out the size of the education effect seems reliant upon three main factors:

First, the setting of the study, where studies conducted outside the US can expect a much weaker effect. This adds weight to the idea that the tolerance enhancing effect of education works to a perhaps major extent through the socialising process of a particular education system, rather than by expanding people’s mental worlds or capacity to appreciate difference.

Second, the operationalisation of political tolerance, where the old critique against the fixed outgroup operationalisation holds up well. That operationalisation of complex theoretical constructs such as political tolerance affects results is no surprise, and this study adds to the argument that researchers must carefully consider the biases introduced by methodological choices.

Third, socioeconomic and –political co-correlates of education and tolerance, mainly economic status and (the in itself complex construct) political knowledge, resources and efficacy. The direction of the effect of an income control variable is a particularly interesting and quite unexpected result. This could itself be a fruitful avenue for further research, in the form of a research synthesis or an in-depth primary study.

It is also noteworthy that the seemingly theoretically sound variables related to personality or psychology did not show any great impact on the education effect size. Perception of threat in particular is a variable often espoused as relevant and with great explanatory power, as is ideology and/or personality, often in the form of so called tendency for right-wing authoritarianism. The results of this study suggest such variables may not have the impact sometimes assumed. One should be cautious about drawing far-reaching conclusions about the relationships of socio-psychological constructs, however, since the interrelatedness of such variables often makes clear distinctions and claims of (non-) relationships problematic.
References


# Appendix A: Included samples

## Table 6. Included samples

<table>
<thead>
<tr>
<th>ID</th>
<th>Identifier</th>
<th>Sample age</th>
<th>Country of sample</th>
<th>Tolerance measure</th>
<th>Tolerance outgroup</th>
<th>n</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bobo &amp; Licari 1989</td>
<td>31</td>
<td>US</td>
<td>Fixed</td>
<td>Mixed</td>
<td>1025</td>
<td>0.20</td>
</tr>
<tr>
<td>2</td>
<td>Djupe &amp; Calfano 2012</td>
<td>5</td>
<td>US</td>
<td>Disliked</td>
<td>Mixed</td>
<td>261</td>
<td>0.02</td>
</tr>
<tr>
<td>3</td>
<td>Duch &amp; Gibson 1992 (A)</td>
<td>27</td>
<td>Non-US</td>
<td>Fixed</td>
<td>Right/Con</td>
<td>589</td>
<td>0.10</td>
</tr>
<tr>
<td>4</td>
<td>Duch &amp; Gibson 1992 (B)</td>
<td>27</td>
<td>Non-US</td>
<td>Fixed</td>
<td>Right/Con</td>
<td>557</td>
<td>0.13</td>
</tr>
<tr>
<td>5</td>
<td>Duch &amp; Gibson 1992 (C)</td>
<td>27</td>
<td>Non-US</td>
<td>Fixed</td>
<td>Right/Con</td>
<td>760</td>
<td>0.01</td>
</tr>
<tr>
<td>6</td>
<td>Duch &amp; Gibson 1992 (D)</td>
<td>27</td>
<td>Non-US</td>
<td>Fixed</td>
<td>Right/Con</td>
<td>679</td>
<td>0.09</td>
</tr>
<tr>
<td>7</td>
<td>Duch &amp; Gibson 1992 (E)</td>
<td>27</td>
<td>Non-US</td>
<td>Fixed</td>
<td>Right/Con</td>
<td>370</td>
<td>0.13</td>
</tr>
<tr>
<td>8</td>
<td>Duch &amp; Gibson 1992 (F)</td>
<td>27</td>
<td>Non-US</td>
<td>Fixed</td>
<td>Right/Con</td>
<td>663</td>
<td>0.09</td>
</tr>
<tr>
<td>9</td>
<td>Duch &amp; Gibson 1992 (G)</td>
<td>27</td>
<td>Non-US</td>
<td>Fixed</td>
<td>Right/Con</td>
<td>569</td>
<td>-0.05</td>
</tr>
<tr>
<td>10</td>
<td>Duch &amp; Gibson 1992 (H)</td>
<td>27</td>
<td>Non-US</td>
<td>Fixed</td>
<td>Right/Con</td>
<td>199</td>
<td>0.17</td>
</tr>
<tr>
<td>11</td>
<td>Duch &amp; Gibson 1992 (I)</td>
<td>27</td>
<td>Non-US</td>
<td>Fixed</td>
<td>Right/Con</td>
<td>650</td>
<td>0.16</td>
</tr>
<tr>
<td>12</td>
<td>Duch &amp; Gibson 1992 (J)</td>
<td>27</td>
<td>Non-US</td>
<td>Fixed</td>
<td>Right/Con</td>
<td>573</td>
<td>0.05</td>
</tr>
<tr>
<td>13</td>
<td>Duch &amp; Gibson 1992 (K)</td>
<td>27</td>
<td>Non-US</td>
<td>Fixed</td>
<td>Right/Con</td>
<td>435</td>
<td>0.01</td>
</tr>
<tr>
<td>14</td>
<td>Duch &amp; Gibson 1992 (L)</td>
<td>27</td>
<td>Non-US</td>
<td>Fixed</td>
<td>Right/Con</td>
<td>484</td>
<td>0.20</td>
</tr>
<tr>
<td>15</td>
<td>Froese, Bader &amp; Smith 2008</td>
<td>17</td>
<td>US</td>
<td>Fixed</td>
<td>Mixed</td>
<td>1023</td>
<td>0.27</td>
</tr>
<tr>
<td>16</td>
<td>Golebiowska 2006</td>
<td>14</td>
<td>Non-US</td>
<td>Fixed</td>
<td>Mixed</td>
<td>.</td>
<td>0.07</td>
</tr>
<tr>
<td>17</td>
<td>Gibson 1987</td>
<td>37</td>
<td>US</td>
<td>Disliked</td>
<td>Right/Con</td>
<td>359</td>
<td>0.06</td>
</tr>
<tr>
<td>18</td>
<td>Gibson 1992</td>
<td>28</td>
<td>US</td>
<td>Disliked</td>
<td>Mixed</td>
<td>1481</td>
<td>0.15</td>
</tr>
<tr>
<td>19</td>
<td>Gibson &amp; Duch 1993 (A)</td>
<td>25</td>
<td>Non-OECD</td>
<td>Disliked</td>
<td>Mixed</td>
<td>478</td>
<td>-0.05</td>
</tr>
<tr>
<td>20</td>
<td>Gibson &amp; Duch 1993 (B)</td>
<td>25</td>
<td>Non-OECD</td>
<td>Disliked</td>
<td>Mixed</td>
<td>1459</td>
<td>-0.05</td>
</tr>
<tr>
<td>21</td>
<td>Karpov 1999 (A)</td>
<td>22</td>
<td>Non-OECD</td>
<td>Fixed</td>
<td>Mixed</td>
<td>1649</td>
<td>0.199</td>
</tr>
<tr>
<td>22</td>
<td>Karpov 1999 (B)</td>
<td>22</td>
<td>US</td>
<td>Fixed</td>
<td>Mixed</td>
<td>1606</td>
<td>0.241</td>
</tr>
<tr>
<td>23</td>
<td>Iglic 2010 (A)</td>
<td>15</td>
<td>Non-US</td>
<td>Fixed</td>
<td>Mixed</td>
<td>1827</td>
<td>0.14</td>
</tr>
<tr>
<td>24</td>
<td>Iglic 2010 (B)</td>
<td>15</td>
<td>Non-US</td>
<td>Fixed</td>
<td>Mixed</td>
<td>1071</td>
<td>0.01</td>
</tr>
<tr>
<td>25</td>
<td>Iglic 2010 (C)</td>
<td>15</td>
<td>Non-US</td>
<td>Fixed</td>
<td>Mixed</td>
<td>1232</td>
<td>0.02</td>
</tr>
<tr>
<td>26</td>
<td>Iglic 2010 (D)</td>
<td>15</td>
<td>Non-US</td>
<td>Fixed</td>
<td>Mixed</td>
<td>992</td>
<td>-0.01</td>
</tr>
<tr>
<td>27</td>
<td>Iglic 2010 (E)</td>
<td>15</td>
<td>Non-US</td>
<td>Fixed</td>
<td>Mixed</td>
<td>1600</td>
<td>-0.05</td>
</tr>
<tr>
<td>28</td>
<td>Iglic 2010 (F)</td>
<td>15</td>
<td>Non-US</td>
<td>Fixed</td>
<td>Mixed</td>
<td>774</td>
<td>-0.22</td>
</tr>
<tr>
<td>29</td>
<td>Iglic 2010 (G)</td>
<td>15</td>
<td>Non-US</td>
<td>Fixed</td>
<td>Mixed</td>
<td>2088</td>
<td>0.01</td>
</tr>
<tr>
<td>30</td>
<td>Iglic 2010 (H)</td>
<td>15</td>
<td>Non-OECD</td>
<td>Fixed</td>
<td>Mixed</td>
<td>720</td>
<td>-0.10</td>
</tr>
<tr>
<td>31</td>
<td>Iglic 2010 (I)</td>
<td>15</td>
<td>Non-OECD</td>
<td>Fixed</td>
<td>Mixed</td>
<td>828</td>
<td>-0.09</td>
</tr>
<tr>
<td>32</td>
<td>Iglic 2010 (J)</td>
<td>15</td>
<td>Non-OECD</td>
<td>Fixed</td>
<td>Mixed</td>
<td>694</td>
<td>-0.04</td>
</tr>
<tr>
<td>33</td>
<td>Gibson 2013</td>
<td>6</td>
<td>US</td>
<td>Disliked</td>
<td>Mixed</td>
<td>3982</td>
<td>0.16</td>
</tr>
<tr>
<td>34</td>
<td>Peffley, Knigge &amp; Hurwitz 2001</td>
<td>21</td>
<td>US</td>
<td>Fixed</td>
<td>Mixed</td>
<td>909</td>
<td>0.15</td>
</tr>
<tr>
<td>35</td>
<td>Rhodes 2012</td>
<td>10</td>
<td>US</td>
<td>Fixed</td>
<td>Mixed</td>
<td>1721</td>
<td>0.19</td>
</tr>
<tr>
<td>36</td>
<td>Lee 2013</td>
<td>3</td>
<td>Non-OECD</td>
<td>Fixed</td>
<td>Mixed</td>
<td>806</td>
<td>-0.04</td>
</tr>
<tr>
<td>37</td>
<td>Tsang 2013</td>
<td>7</td>
<td>Non-OECD</td>
<td>Fixed</td>
<td>Mixed</td>
<td>836</td>
<td>0.13</td>
</tr>
<tr>
<td>38</td>
<td>Cigler &amp; Joslyn 2002 (A)</td>
<td>38</td>
<td>US</td>
<td>Fixed</td>
<td>Mixed</td>
<td>5986</td>
<td>0.28</td>
</tr>
<tr>
<td>39</td>
<td>Cigler &amp; Joslyn 2002 (B)</td>
<td>25</td>
<td>US</td>
<td>Fixed</td>
<td>Mixed</td>
<td>1556</td>
<td>0.20</td>
</tr>
<tr>
<td>40</td>
<td>Golebiowska 1995</td>
<td>27</td>
<td>US</td>
<td>Fixed</td>
<td>Mixed</td>
<td>480</td>
<td>0.30</td>
</tr>
<tr>
<td>41</td>
<td>Seligson &amp; Caspi 1983</td>
<td>32</td>
<td>Non-OECD</td>
<td>Fixed</td>
<td>Mixed</td>
<td>490</td>
<td>0.14</td>
</tr>
<tr>
<td>42</td>
<td>Petersen &amp; al 2011</td>
<td>9</td>
<td>Non-US</td>
<td>Fixed</td>
<td>Mixed</td>
<td>1479</td>
<td>0.10</td>
</tr>
<tr>
<td>43</td>
<td>Weldon 2006</td>
<td>18</td>
<td>Non-US</td>
<td>Disliked</td>
<td>Mixed</td>
<td>5730</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Appendix B: Notes on coding and data manipulation

Table 7. Notes on coding and data manipulation

<table>
<thead>
<tr>
<th>Preferred sample</th>
<th>When several models were fitted to a single sample, the most comprehensive model was marked as “Preferred”, based on a qualitative judgement of:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a) Judgement of original author/s.</td>
</tr>
<tr>
<td></td>
<td>b) Most comprehensive model, in terms of validity of tolerance measure, control variables included, and sample size.</td>
</tr>
<tr>
<td></td>
<td>c) When one sample was used for several measurements of tolerance, one was chosen in order of preference: 1) Disliked group, 2) Fixed group, 3) Civil liberties, 4) Other.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>Age of sample. Calculated, in order of preference, as</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a) 2015 – Year of sampling, and for multiple year samples, using one year if year of sample is controlled for, and mean year if not (rounded down).</td>
</tr>
<tr>
<td></td>
<td>b) 2015 – Year of published article, when sample year was not available.</td>
</tr>
</tbody>
</table>

| Country          | If a sample was drawn from a single country, it was coded as only-US, non-US, and, if applicable, also non-OECD (as a measure of developed versus less developed countries). If it was drawn from several countries, it was coded according to the included countries: Non-US and, if applicable to any country in the sample, non-OECD. |

| Political tolerance measure | This is a qualitative judgement based on the nature of the political tolerance measure in a model. If it was based on questions pertaining to the rights of pre-selected outgroups, in the vein of Stouffer/GSS, it was coded as “Fixed group”. If it was based on respondent selected outgroups, following Sullivan, it was coded as “Disliked group”. If it was based on the rights of all citizens, it was coded as “Civil liberties”. If another measure was used, it was coded as “Other”. |
If it was based on a fixed group, but controlled for dislike (i.e., only those who found the pre-selected group objectionable are included in the sample), it was a qualitative judgement based on the phrasing of the “dislike” component. Weak dislike control (e.g., everyone except those who actively favour the group are included) were labelled as “fixed group”, while those with a strong dislike control (e.g., only those who actively object to the group are included) were labelled as “disliked group”.

<table>
<thead>
<tr>
<th>$\beta$, Effect size</th>
<th>These were gathered directly from relevant regression output. If $\beta$ was not reported, it was calculated when $b$, $s_{edu}$ and $s_{tol}$ were available, according to: $\beta = b * \frac{s_{edu}}{s_{tol}}$. In some cases, path analysis Pearson correlation was used as a substitute, following the relative interchangeability of the measurements following Peterson &amp; Brown (2005).</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>Preferably taken directly from regression output, otherwise from in-text account of sample. In one case (Golebiowska, 2006) imputed as mean of $n$ sizes, due to missing data.</td>
</tr>
<tr>
<td><strong>Controls</strong></td>
<td>Coded 1 if model includes one, or several, controls of this type.</td>
</tr>
</tbody>
</table>