On the Measurement of Social Belonging and its Connection to Migration Background

by

Nathan Dale Roberson
B.A., University of Colorado at Boulder, 2011
M.A., Central European University, 2012

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Measurement, Evaluation, and Research Methodology)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

March 2020

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

On the Measurement of Social Belonging and its Connection to Migration Background

submitted by Nathan Dale Roberson in partial fulfillment of the requirements for the degree of Doctor of Philosophy

in Measurement, Evaluation, and Research Methodology

Examining Committee:

Dr. Bruno D. Zumbo, Professor and Distinguished Scholar, Faculty of Education, UBC
Supervisor

Dr. Anne Gadermann, Assistant Professor, School of Population and Public Health, UBC
Supervisory Committee Member

Dr. Amery Wu, Associate Professor, Faculty of Education, UBC
Supervisory Committee Member

Dr. Skye Barbic, Assistant Professor, Faculty of Medicine, UBC
University Examiner

Dr. Kimberly Schonert-Reichl, Professor, Faculty of Education, UBC
University Examiner

Abstract

Social belonging is a central human need, and one's immigration background is an important factor when considering how we make sense of the measurement of social belonging. Using the Programme for International Student Assessment (PISA), I situate the measurement of social belonging in schools within an international comparative context. Through multi-method Differential Item Functioning (DIF) studies, and the use of a multi-level validation processes, the evidence presented here suggests that individual-level characteristics of immigration background and country-level characteristics such as national multicultural integration policy are valuable explanatory variables to understand the ecological validity of social belonging in schools. By reading the data through a "diffractive" methodology, traditional psychometric evidence can be taken up as a part of contemporary, situative theories of education. These findings have important implications for the fields of educational measurement, political science, and immigration studies alike. These (inter-)disciplinary contributions likewise advance a novel approach to measurement that ties psychometrics to an agenda of democratic liberalism in an age of deep diversity.

Lay Summary

Social belonging is a central human need. I investigate to what extent immigration background is an important explanatory variable of one's sense of belonging. Using the Programme for International Student Assessment (PISA), I situate the measurement of social belonging in schools within an international comparative context. The evidence presented here suggests that individual-level characteristics of immigration background and country-level characteristics such as national multicultural integration policy are valuable explanatory variables in the ecological validity argument. Then, I have the goal of exploring psychometric evidence as a part of contemporary theories of education and political science. Finally, I explore the political implications of this evidence in an age of deepening diversity.

Preface

Chapter 3. A version of this material is published as:

Roberson, N. D., & Zumbo, B. D. (2019). Migration Background in PISA's Measure of Social Belonging: Using a Diffractive Lens to Interpret Multi-Method DIF Studies. *International Journal of Testing*, 19(4), 363-389. https://doi.org/10.1080/15305058.2019.1632316

For this chapter I was the lead investigator and author including conducting all analyses, writing the manuscript, and submission and revisions. Dr. Zumbo provided invaluable guidance and feedback.

Data for the dissertation comes from the Organization of Economic Cooperation and Development is publicly available of the PISA website. As such no BREB was needed. Data was published in December, 2016.

This dissertation is an original intellectual product of the author, Nathan Dale Roberson.

Table of Contents

List of Tables

List of Figures

Acknowledgements

For the duration of my studies I have lived, worked, and played on the traditional, ancestral, and unceded territories of the Musqueam peoples. I consider this privilege a joy and a responsibility.

This work represents a culminated effort of thoughts and experiences shared with countless persons and across many times. Thus, to all of my friends and strangers who have encouraged me and kindly accepted my interrogations about immigration, measurement, policy, and general thoughts – Thank you!

In particular, I wish to thank Dr. Beverly Webber, Dr. Dayna Matthew, and Dr. Jane Menken who counselled me during my time at CU Boulder. Your support and encouragement on the path to become a scholar has truly transformed the trajectory of my life. Likewise, your examples of using scholarship to advance social justice is a continued inspiration for all current and new scholars.

I wish to thank Dr. Alexander Astrov at Central European University who nurtured me into the Continental European Philosopher I consider myself today. Likewise, I am grateful for your example of transgressing disciplinary boundaries and your lived experience of how to continue to advance liberal democratic values while living under oppressive regimes. Particularly as we continue to struggle for a more just and compassionate world in an age where democracy is in recession and immigrants are increasingly villainized.

At the University of British Columbia, I am grateful for the incredible support of the MERM program and fellow students. In particular Dr. Amery Wu and Dr. Anita Hubley who steered me in my early years in Canada. Then to my "academic father", Bruno D. Zumbo, who has truly struggled through the complexities of thinking beyond our present thought, and how to reconceive of our entire field of study as we await an epistemological revolution. Thank you! I am happy to be a part of your strange family, and I take the "adoption" seriously as a motivator and blessing.

For my colleagues as a part of my work at the Centre for Teaching, Learning, and Technology – I wish to express my gratitude for your example of how to always keep scholarship tied to practice.

On a personal note, I owe a debt of gratitude to my family who collectively redefined my world-view; in particular my sister, Amy, who has carried my burdens alongside me in more ways than can be expressed; and my FARMily – who have nurtured my body and soul.

And of course, Matthew, I consider this dissertation the fulfillment of a specific promise to you, and the symbol of a continued promise. For me, belonging is forever defined in relationship to you.

Dedication

For all of those who lives have been touched by immigration.

Chapter 1: Measuring Social Belonging in the Age of Migration: Methodological and Political Considerations

We are living in an age of unprecedented internal migration and international immigration. According to the International Organization for Migration (IOM), the last global count of migration in 2017 recorded 258 million immigrants globally, which represents 3.4% of the global population, compared to 2.8% in 2000, and 2.3% in 1980 (IOM, 2018). Across nearly all domains: students, labour migrants, refugees, and irregular migrants, the number of persons crossing international boundaries continues to increase in an ever-globalized world. Within respective countries, roughly 750 million persons are considered "internal" migrants (United Nations, 2018), meaning that collectively 1 in 7 persons is a migrant across the globe. Meanwhile, more than half of all international immigrants live in just ten countries; one of which is Canada (United Nations, Population Division, Department of Economic and Social Affairs (UN DESA), 2017). Along with increased urbanization, immigration increasingly is contributing to new and diverse social fabrics in our societies. In Canada, the federal government has set forth an ambition immigration plan to welcome nearly 1 million immigrants between 2018-2020 alone (IRCC, 2017). In a country of roughly 35 million people, this represents a sizeable portion of the overall population, some of whom are existing immigrants. While much of the formal immigration plans are driven by targets for economic integration, social cohesion remains paramount objective. With ever more diverse societies, our ability to maintain and foster a sense of communal belonging is

increasingly important in order to maintain and increase levels of health, happiness, education, and other outcomes discussed throughout this dissertation. A part of this challenge is actually measuring levels of social belonging to support interventions.

My dissertation takes an empirical investigation into the measurement of social belonging in schools and its connection to migration background. However, while my empirical studies examine a particular measure of social belonging in schools, I am broadly interested in social belonging at large, and I take for granted that students' sense of belonging in school is one instantiation of overall social belonging. As such, when I am exploring broader theoretical discussions and their implications I use the shorthand of "social belonging", and when referring to my empirical investigation I make an effort so specify "social belonging in schools". The studies presented herein are empirical in nature, they are grounded in, and build upon, existing theories of validity and validation. In particular, I stress the importance of considering one's larger ecological setting as a part of validity claims. More specifically, I see this work as an extension of the "integrated framework" of evidence by including sociological considerations. Indeed, "the field is finally moving toward a state in which useful procedures will be available to the practitioner trying to develop a test that meets some standard for test development and test validation" (Lissitz, 2009, p. 5). To be clear, throughout my dissertation, I adopt the term, "ecological validity" to describe the importance of considering sociological variables and context as a part of the validity

argument, but I am not proposing it as a "type" of validity.[1] Indeed, consistent with Fox, "from an ecological perspective, individuals do not exists as isolated units; rather they are dynamic, socially embedded, and defined by a network of relationships" (2003, p. 22). That is, to say "ecological validity", is to draw an explicit recognition that all assessments, and likewise all validity arguments, exist within a certain ecological setting.

It my view, shared with others, that consideration of immigration and the contexts therein should be considered in the development of sociological assessments. This shift towards consideration of various procedures and methods begins to account for the myriad of ways in which validation evidence can be collected to support an argument for validity. Building on the work of Messick, Kane advances an argumentative approach to validity where "the need for validation derives from the scientific and social requirement that public claims and decisions be justified" (Kane, 2006, p. 17). This focus on interpretations and meanings of measurement scores to justify decisions reflects implicit and explicit values inextricably linked to their (un)intended consequences.

As Zumbo (2009) sees it, one of the largest problems with more traditional approaches to validation evidence is their tendency to adopt a context-free approach to validation evidence. In particular, Zumbo critiques the deductive-nomological network adopted by Cronbach and Meehl (1955) and extended in Messick's work.

[1] I assume and adopt the notion of a "unified" framework of validity consistent with contemporary psychometric literature, rather than the triarchic view or others.

> Validity is not some sort of super-knowledge of the phenomenon one wishes to measure, such as the embodied meta-theoretical views of Messick, Cronbach and Meehl, and myself, but simply more knowledge: knowledge of causes" (Zumbo, 2009, p. 73).

Hubley and Zumbo (2011) also argue that consideration of consequences is important in test interpretation.

> It is important to reflect upon and understand the values that underlie our constructs, measures, and measurement because they impact the meaning of the test scores, the relevance and utility of inferences made with different samples, contexts, and time periods, and the consequences of test use (Hubley & Zumbo, 2011, p. 223).

In the case of social belonging, particularly with a sample in part made up of immigrants, our social values and integration policies become a part of the wider net that begins to delimit the construct since integration policy defines how immigrants are extended formal belonging. For Kane, the argumentative approach to validity consists of two arguments: an interpretive argument and a validity argument. The interpretive argument "proposes interpretations and uses of test results by laying out the network of inferences and assumptions leading from the observed performances to the conclusions and decisions" (Kane, 2006, p. 23). Then, the validity argument is an evaluation of the interpretative argument. Accordingly the argument "should make clear . . . the construction of reality and the value weightings implicit in a test and its application" (Cronbach, 1988, p. 5). Thus to claim that an interpretation is valid "is to claim that the interpretative argument is coherent, that its inferences are reasonable, and that its assumptions are plausible" (Kane, 2006, p. 23). Overall he is advancing an "approach to validity rather than a type of validity" (Kane, 1992, p. 534). In establishing an approach

towards validity, the argument is not binary (valid or not), but rather it is a matter of degree. Indeed, it is precisely this theoretical shift, that I am advancing in this dissertation by presenting a case for the validity of our use and interpretation of PISA's measure of social belonging in schools with respect to migration background.

In addition to changing conceptions of validity, statistical approaches are also changing. In this dissertation, I use contemporary statistical approaches to analyze my data, and I connect these results to evolving measurement theory. Revised methods and approaches to validity then have broader implications; including policy implications. As such, I trace the importance of these methodological contributions in a broader context of philosophy of science.

A version of Chapter 3, entitled "Migration Background in PISA's Measure of Social Belonging: Using A Diffractive Lens to Interpret Multi-Method DIF Studies" has already been published in the International Journal of Testing. This chapter uses a diffractive lens to investigate DIF using Ordinal Regression and the "Alignment" Method. Here I introduce how one's migration background is an important consideration for DIF, particularly with sociological scales. This chapter highlights how multi-method studies should not be conducted merely to look for corroborating evidence, but it is also important to read between the contradictions and tensions of the results. In this chapter, the primary investigation is to understand DIF as it relates to migration background. As such, while treated independently of one another, countries are considered more or less "interchangeable." That is, the alignment method and OLR ask the question if, for example, foreign-born persons can be considered as "statistically

equivalent" on the PISA measure of belonging when compared to native-born persons. This is in fact a common type of assumption that is frequently made in large-scale assessments as we compare the math performance of females in Canada and Germany thereby assuming that females are interchangeable in both countries. Knowing that this is an assumption, I then seek to investigate the multi-level nature of this data as a part of the subsequent research chapter.

Chapter 4 is entitled "Multicultural Integration Policy as an Explanatory Factor of Immigrant Social Belonging: Multilevel Evidence for a Multilevel Construct." This research builds upon the research of multi-level construct validation and provides evidence for the use of multi-level methods in the case of PISA's belonging scale; with consideration for migration background. PISA, and other large-scale assessments, are frequently used to rank countries to help policy makers (re)shape respective policy frameworks to achieve important goals. In this sense, it is worth noting how policies are treated as a type of intervention in their respective countries; acting as a political lever to encourage the development and formation of their denizens. Looking at the same construct, PISA's measure of belonging, I consider how this construct demands multi-level consideration. This particular measure is collected from individuals about their personal experience in schools. These data are then aggregated to understand country-level characteristics. As such, these data actually represent a three-level construct. Moreover, this research is the first paper that I am aware of that introduces country-level policy characteristics using the "Migration Integration Policy Index" (MIPEX) as a part of the validity argument of a sociological scale like sense of belonging. This

research has important implications for the field of psychometrics by extending the ecological framework of item-response to include the policy domain. Likewise, as a policy piece, this research extends the development and use of policy indices to be of use in the measurement community. In general, descriptive indices of policies are the end goal of researchers that can be used to contextualize, rank, and compare countries to one another. However, I extend the use of these indices to be used in multi-level models as a part of psychometric evidence. As such, I seek to create a bridge between the fields of political science and psychometrics, and I provide an empirical example of how this research fits together in an ecological sense as proposed by Bronfenbrenner and others.

Finally, chapter 5 provides a conclusion that further extends the diffractive method as introduced In chapter 3 to chapter 4. Here I highlight how these closely related research projects fit together within a larger body of research. Likewise, I further outline future directions of this research. Most importantly, chapter 5 provides a synthesis of the philosophical implications of the work that emerged throughout the research process. Namely, I make the argument that the collective body of work presented here provides a justification to place my empirical, quantitative work within the educational context of "situative theory" and how other scholars engaging in ecological validity can also be read in this light. It is worth noting a bit of the nuance about the way in which I take up an ecological model. In the traditional "ecological framework" of Bronfenbrenner and others, he draws attention to the environmental systems with which an individual *interacts* to shape development. However, while I

accept and explore this theory, I also seek to frame validity here within a theoretical lens of situative theory (Nasir & Saxe, 2003); which highlights how knowledge, thinking, and learning are situated/located within experience. As such, situative theory is more than an interaction of an individual and environmental systems and it assumes that the ecological setting cannot be separated from the individual experience.

This dissertation is a "manuscript" model. As such, at times the literature cited is intentionally redundant throughout the chapters. Likewise, although chapter 3 has already been accepted for publication, and chapter 4 has been submitted, I make some slight modifications to their presentation here for the sake of a cohesive dissertation work. This includes adding in additional details or thoughts that were not warranted for publication but are useful as a part of an educative piece.

It is important to note at the outset, that consistent with my prior academic training and many of the scholars cited throughout my research, I accept that I make certain assumptions as a scholar of International Relations. Namely, like Bronfenbrenner, Bloemraad, Herztman, Taylor, Zumbo and others referenced throughout my piece, I make the assumption that national characteristics must play a role in formulating the context in which we live and develop. I assume, and ultimately test in chapter 4, that countries are important. This seems perhaps particularly obvious when considering the connection of belonging to migration background. Migration background is, in many ways, an artificial construct. It is "performative" in the way that Judith Butler defines it in reference to gender. Namely that the boundaries of "foreign-ness" and "native-ness" are artificial and non-existent and simultaneously perceived,

reified, and enforced in the Althussian and Foucauldian sense through our use of categories that become "policed."

Yet, just as migration background is performative, it is defined in a relational way. That is, for example, I am considered "foreign" here in Canada (even with many shared histories and cultures), but I am "native" inside the United States. These definitions are accordingly "place-based" and dependent upon the particular place of birth of oneself and their parents, and their current place of references. It is not lost on me that there are many places *within* the United States that I would, in many ways, be much more "foreign" than the majority of places in Canada. This preceding claim highlights the way in which national-level identity is also constructed, or is an "imagined identity" as defined by Benedict Anderson; that the definition of a "Canadian" or "US citizen" is a mere act of collective will to believe such a thing exists. Yet, consistent with Butler, these problematic labels carry with them a gravity that is experienced upon our physical and mental bodies. Despite my "successful" integration into a Canadian context, including being labeled a "permanent resident" I am still unable to vote, I am still perceived as an outsider, and there is much about the "hidden curriculum" of Canadian life that I am still learning; and I am still cast as suspect for not knowing. Consistent with the works of Gloria Anzaldúa, I am intentionally highlighting in this introduction the ways in which our physical/mental bodies, how my physical body, is the place in which theories of politics and theories of measurement are played out. Moreover, how, paraphrasing Anzaldúa, "the bridge called my back" is the site in

which the larger theories are contested and simultaneously the generative site of new theories and epistemologies.

The research presented in the dissertation is quantitative in nature, and it is presented in a formal way embedded within the fields of psychometrics and political science. However, I would be bereft if I did not take the opening introduction to also acknowledge that it is incredibly personal. Especially for those of us in North America, we almost all have a connection as immigrants or in relation to immigrants. Nowhere is this truer than as I am writing this piece on the unceded territories of the Musqueam peoples at the University of British Columbia. I am simultaneously a guest on this land as a student at UBC, and I am a white-colonizer on this land as a settler living here by permission granted from the Canadian and formerly British colonial power. All of these tensions are incredibly relevant, although not discussed, as I explore what does it mean "to belong." Likewise, while mostly absent in this formal presentation, these are teaching and learnings that are also a part of my educational experience at UBC living alongside the Coast Salish nations.

The following research presentation is divided into five chapters. Chapter 1, this chapter, provided an introduction to the main bodies of research work. Chapter 2 is a background chapter which provides a more thorough discussion about the construct of social belonging and a discussion about the methods employed in this dissertation. Then, chapters 3 and 4 represent the main bodies of empirical research. Finally, Chapter 5 provides concluding thoughts and connects the empirical research to a broader philosophy of sciences and political theory.

Chapter 2: Background

In this chapter I situate my research projects within the larger fields of measurement, sociological theories of belonging, and political theories of immigration and policy. Namely, this chapter explores key concepts that are central throughout this research project: multi-level validation, measurement invariance, social belonging as a construct, the existing psychometric evidence of belonging, and a discussion of immigration background as it relates to policy.

Validation and Multilevel Consideration

As a part of a cross cultural investigation of social belonging, one also needs to be cognizant of the multilevel nature of the analysis. This is to say the contextual factors of different migrant groups vary tremendously from country to country. "Multilevel validation methods aim to provide a strong form of construct validity; that is, the evidence should provide an explanation for the observed variation in test scores" (Zumbo, Liu, Wu, Forer, & Shear, 2017, p. 342). Multilevel validation is more than analyzing multilevel data, but offers an explanatory model about the construct under investigation. This is to say, more than simply investigating measurement invariance of social belonging based upon one's migration background, I seek to understand the cross-national variation in social belonging and it's connection to immigration background using an *explanatory model*.

Large-scale international assessments are explicitly designed to provide feedback at an aggregate (school/region/country) level and they are not intended to provide any feedback for the individual respondents. Rather they are used to inform policy systems.

Within the context of educational settings most constructs are inherently multilevel and methodological consideration should be assumed (Klein, Dansereau, & Hall, 1994). Given that multilevel constructs are used to shape policy in schools and countries for all individuals in a target population in a "high stakes" context, it is important that adequate conceptual and methodological thought is given to the evaluation of these measures.

Questions about belonging are asked of individuals, who consider their "group-belonging" when understanding one's individual sense of belonging (G. L. Cohen & Garcia, 2005). In short, this requires an assessment of the level of theory in order to explain data variability in multilevel validation (Zumbo & Forer, 2011). Since the level of measurement of belonging is at the student-level within schools, and the inferences are being made about groups within countries, there is enormous potential for spurious claims. To provide more accurate findings, analysis needs to properly consider the structure of the data. More specifically, we ought to avoid the "atomistic fallacy", where unjust inferences are drawn at the aggregate level from individual level data; and the ecological fallacy, where unjust inferences are made about individuals from aggregate level data (Bliese, 2000).

Within the context of multilevel research, Chen, Matthieu, Bliese, Yammarino, and Dansereau (2004) propose six compositional models. The selected-score model uses a single, individual score to define the group-level construct. Second, the summary index model uses a statistic (such as mean or sum score) of individual scores to define the group. Third, the consensus model which measures within group agreement using

items about another level (e.g. individuals rate their own group motivation). Fourth, the referent-shift consensus model which measures within group agreement about items pertaining to the group. Fifth, the dispersion model measures within-group heterogeneity (e.g. differing skills in a team). Then the sixth, aggregate properties model measures constructs directly at the group-level (e.g. supervisor rates an entire team's productivity).

Once the framework a multilevel construct is understood, one can investigate the patterns of variability across levels to draw conclusions. A developed approach is the within- and between- group analysis of variability (WABA: Da Costa & Araújo, 2012; Dansereau & Yammarino, 2000; Forer & Zumbo, 2011). Groups can be composed of homogenous, independent, or interdependent members; referred to as wholes, equivocal, and parts, respectively. If group members are homogenous, a single value can describe the entirety of the group; which assumes that only between-group variation exists in the data/construct under investigation. Conversely, with independent members, all variation is assumed to be at the individual level. Lastly, the parts view suggests there is individual variation, but it is dependent upon one's group context.

To make inferences using WABA, Dansereau and Yammarino (2000) propose two statistics to be used: E-ratio to assess practical significance and the F-ratio to assess statistical significance. One should first confirm the F-ratio before assessing the inference of the E-ratio. The F-ratio is the typical Mean Square(between)/Mean Square(within) when a "wholes" view is implied, and is the inverse ratio for "parts"

view. The E-ratio is the between-group eta correlation divided by the within-group eta correlation. When within- and between- group variation is equal, the E-ratio is 1[2]. The E-ratio has a lower bound of 0 (no variation between groups), and no upper bound. For both ratios, one or more predetermined critical values need to be set before making inferences (Dansereau, Alutto, & Yammarino, 1984). The 15° test divides the overall variance into three equal parts, while the 30° test creates equal intervals in terms of angles and is the more conservative test. The 30° test requires within- or between-group variation to exceed 75% of the total before an inference about wholes or parts may be made. Social belonging is an individual experience, but also has a group component to it (discussed further below). A proper analysis of social belonging requires that we adopt a multi-level, explanatory approach, particularly in the context of cross-cultural comparison.

Measurement Invariance

Differential Item Functioning (DIF) is the expression of "some characteristic of the test item and/or testing situation that is not relevant to the underlying ability of interest and hence test purpose" (Zumbo et al., 2015, p. 137), and accordingly the item (or scale) may be measuring different abilities across sub-groups. The presence of DIF, suggests that items do not adequately capture the construct under investigation equally across groups. When DIF is present, one's ability to draw conclusions is compromised. As such, inferences cannot be attributed to just a participants performance or latent

[2] This is equivalent to Cohen's f effect size statistic.

ability, but may be due to another measurement artifact (F. Chen & West, 2008). Said another way, DIF can be expressed either when one thinks they are investigating the same construct, but the construct may actually be different, although likely closely related. Or, when one is measuring something in excess to the construct at hand, which is called construct irrelevant variance. More broadly, DIF is central to the practice of validation itself (Gómez-Benito, Sireci, & Padilla, 2018) since it compromises our ability to interpret a test score at all, and accordingly draw inferences.

One cannot properly study the impact of particular measures if DIF is present (Holland & Thayer, 1988). DIF, bias, and impact of items should be accordingly be considered in the context of group comparisons, since findings become specific to the groups under consideration (A. D. Wu, Liu, Stone, Zou, & Zumbo, 2017). In short, DIF is present when differences between groups are evident among respondents with equal levels of the construct being measured (Wu et al., 2017).

It is necessary to ensure measurement invariance to make cross-cultural comparisons on psychological constructs (F. Chen, 2007). Yet, in practice it can be an unrealistic goal. The alignment method was developed precisely to address this challenge. Other approaches are used to assess measurement equivalence, such as multiple group confirmatory factor analysis (MGCFA: Bollen, 1989; Steenkamp & Baumgartner, 1998), item response theory analysis (IRT: Jilke, Meuleman, & Van de Walle, 2015), DIF-via logistic regression (Zumbo, 1999), and latent class analysis (LCA: Kankaraš, Vermunt, & Moors, 2011). The most popular approach is likely MGCFA and it is frequently used in political science to measure cross-country equivalence of values,

like democracy and welfare (Ariely & Davidov, 2011), social and political trust (Freitag & Bauer, 2013), and national identity (Davidov, 2009).

Using MGCFA, it is recommended that one should test for invariance, such as configural/form invariance (items load onto the same factors across heterogeneous groups); factor-level/metric invariance (where the loadings are roughly equal); and the intercept/scalar level (where the scores on the construct have the same unit of measurement and the same origin). Scalar invariance is typically needed to compare latent means. Then, if the factor loadings are equal, metric equivalence is said to be met; where a one-unit increase would mean the same for all groups. "Scalar variance" is the highest level, which constrains the factor loadings and intercepts to be equal across groups. This implies that respondents with the same value on the latent construct have the same expected response on the scale, irrespective of their group (Davidov et al., 2015). Steenkamp and Baumgartner (1998) argue that configural and metrics invariance are not enough to ensure valid comparisons of latent means across groups therefore, scalar invariance is necessary.

The difficulty of cross-national comparisons on social indicators is well documented. Are differences in mean-values "true" differences or an artifact of tool development/implementation in a new context? Adam (2008) investigated cross-cultural social capital using the European Values Survey (EVS) and the European Social Survey (ESS), and found that even within the same country in the same year, mean values of generalized trust differed by groups. They attribute this to differentiated scaling of the ESS and wording effects, and they find that the overall country ranking

remains stable. Freitag and Bauer (2013) used MGCFA and found evidence that social trust was stable across cultural backgrounds. Using a comparison of French, Italian, and German communities in Switzerland, they found that diversity of ethnic background does not constitute a barrier for analysis of social trust at the sub-national level. Underlying this entire dissertation is the question, "Can people with different migration backgrounds be compared to one another?", or said another way, "Is there evidence of DIF across migration backgrounds?" To test for DIF requires a psychometric analysis of the items in question. To date, there is no psychometric evidence that people with different immigration backgrounds are at all comparable statistically in terms of their social belongingness. Meanwhile, policy makers are increasingly invested in projects of acculturation, thus the need to assess measurement invariance across migration background is increasingly important.

The Construct of Belonging

The notion of belonging emerges from the field of anthropology which posits that natural groups are characteristic of all human groups (Coon, 1946), including the famous "Robbers Cave Experiment" in social psychology / social anthropology which showed that people randomly assigned to certain groups formed strong loyalties and group attachments (Sherif, Harvey, Hood, Sherif, & White, 1988). Related, we have evidence that social contact can help replace established intergroup prejudices (Wilder & Thompson, 1980), and that people have a natural impulse towards forming positive attachments over negative oppositional patterns (Orbell, van de Kragt, & Dawes, 1988). This is based on the so-called "contact hypothesis" that posits that interaction across

different groups should reduce hostilities, which stands in contrast to economic-resource models which suggests that conflict emerges when different groups fight for the same resources (Harell & Stolle, 2010; Tajfel, 1981). Harell and Stolle (2010) make clear that these models are insufficient because they fail to account for intergroup contact, and these models assume that geographical diversity at the state or province capture social interaction. Likewise, some scholars have investigated belonging out of a concern that its omission results in negative outcomes; even suicide (Spaulding, Simpson, & Durkheim, 1997). Belonging is so central that it even emerges within political theory. Morgenthau (1962) argues that the pursuit of power is balanced by the pursuit of love, but that both represent an effort to escape loneliness. According to this theory, in order to manage the threat of isolation, one must break down barriers between people, where love aspires to a mutual dissolution of boundaries leading to a new whole, and power is a unilateral overcoming of boundaries.

One challenge to analyzing the construct of social belonging is the relative lack of consensus of defining the term in contrast to its many related terms. Lee and Robbins (1998) define "social connectedness" as an internal sense of belonging that is "the subjective awareness of being in close relationship with the social world." Social connectedness is more than just contact, but also is concerned with the quality and quantity of intimate connections (Reis & Franks, 1994). Lee and Robbins (1995) propose that the need for social belonging is expressed early in one's life and that people satisfy this need through their identification and participation in the broader social world. They find that high levels of connectedness supports cognitive processing and improves

an individual's capacity to self-manage their needs and emotions (Lee & Robbins, 1998). More than just a process to manage one's needs, belonging is a dimension of personal identity. People have personal/intimate notions of self, social/relational aspects of self, and collective ideas of self, and that people are more likely to agree with and receive a personal amount of gratification from the successes of their "in-group" (Cacioppo & Patrick, 2009). As such, the belongingness hypothesis states that humans have a minimum need for lasting, positive, and significant interpersonal relationships dependent upon two criteria: a need for frequent and affectively positive interactions with at least a few people and second, these interactions should occur in a stable and lasting framework of concern for one another's welfare (Baumeister & Leary, 1995).

The problem of social exclusion has been frequently taken up in psychology and sociology. One of the most documented aspects of social exclusions concerns issues of race and racism. Walton and Cohen (2007) document how black students, as a stigmatized group, are often more sensitive to race based rejection, which can contribute to race based disparities in achievement. Walton and Cohen (2011) use an intervention to lessen the perception of threat experienced by African-Americans by framing social adversity as a common and transient experience of university. They build upon others' findings that social belonging, defined as "a sense of having a positive relationship with others," is a central human need (Baumeister & Leary, 1995), and that the lack of belonging, or exclusion, harms people in a number of ways including subjective well-being (Lyubomirsky, Sheldon, & Schkade, 2005), intellectual achievement (Walton & Cohen, 2007), health and immune functioning (Berkman &

Syme, 1979), and self-control (Baumeister, Twenge, & Nuss, 2002). Walton and Cohen saw that by reframing isolation and exclusion as a temporary phenomenon with African-American participants, the minority students halved the "achievement gap" and had fewer doctor visits, improved reported health and well-being in a 3-year post-intervention (2011). Their findings affirm that the perception of social belonging has broad impacts and suggests that social belonging can be regarded as a psychological lever within targeted interventions.

Mendoza-Denton et al. (2002) also studied race-based rejection sensitivity among college students and found declines in grades as a result of race-based rejection over a 2-3 year period, while students with positive race-related experiences have increased feelings of belonging. Race-based sensitivity is a cognitive-affective process where people anxiously expect, readily perceive, and intensely react to occurrences of rejection. Their research confirms findings that anxious expectations of race-based rejections strain relationships and undermine people's confidence in the fairness and legitimacy of institutions, as such institutions need to move beyond diversity purely in numerical terms, but must ensure that members of different groups actually feel that they belong, are accepted, and have trust in institutions (G. L. Cohen, Steele, & Ross, 1999). These findings are consistent with other research suggesting that awareness of one's belonging to a stigmatized group, such as race, religion, gender, ethnicity, or sexual orientation, in a particular value domain results in decreased performance in that domain (Steele, 1997; Steele & Aronson, 1995). Steele (1997) theorizes that school success, and arguably other success, requires identification with the school and its

subdomains and can be compromised through societal pressure; those who belong to a negatively stereotyped group have an added burden of "stereotype threat" from others' judgements that their actions will negatively reinforce stereotypes of their own group. His work shows that women and African-Americans are particularly susceptible to the impacts of stereotype threat, which may result in dis-identification within a domain leading to reduced sustained motivation in that domain.

Mendoza-Denton et al. (2002) tested the generalizability of the stigma-consciousness construct with gays and lesbians, men and women, and multiple racial groups. Theoretically, one can imagine that stigma-consciousness might apply to any group that is frequently judged on stereotypes, particularly if they are visible minorities, but that it may not manifest in the same way for all groups. Pinel (1999) analyzed the applicability of the stigma consciousness questionnaire (SCQ) and found that women, gays and lesbians, and blacks all had significant amounts of reported stigma consciousness. Excessive concern about stigma consciousness or one's stereotype is a barrier to moving beyond it (Pinel, 1999).

For already marginalized and stigmatized individuals, the question of acceptance and belonging is considered the "central feature" of one's life (Goffman, 1986). Similarly, for individuals of a stigmatized group(s), the question of personal belonging is part and parcel with the question of group belonging at large (G. L. Cohen & Garcia, 2005). This is sometimes referred to as "collective threat" when one fears that their in-group belonging might reinforce negative stereotypes of that group. Since people's identities and self-worth are influenced by their groups, peoples' thoughts,

feelings, and actions are impacted by the objective outcomes of their group (G. L. Cohen & Garcia, 2005). This research suggests that the threat of a fellow group member confirming a stereotype may be as acute as personally confirming a negative stereotype. In fact, Bernhardt et al. (1998) document how peoples' self-esteem may increase when fellow members succeed, and may also be evident through changes in one's hormone levels. This is in part because self-esteem depends upon recognition and respect for one's culture (Parekh, 2006).

Belongingness has also been studied frequently in the context of intimate and often romantic relationships (Aron, Aron, Tudor, & Nelson, 1991). Evidence suggests that those suffering from marital problems such as rejection, separation, or divorce, experience greater immune system problems as a result of the exclusion (Kiecolt-Glaser et al., 1987). Berkman and Syme (1979) found that respondents with social ties (marriage, contacts with close friends and relatives, church membership, and group associations) had lower mortality rates than those without such ties. Similarly, research suggests that veterans with a strong support network and sense of belonging experience reduced symptoms of Post-Traumatic Stress Disorder (PTSD: Solomon, Waysman, & Mikulincer, 1990). The need to belong is such a strong motivator that in some communities individuals will engage in risky or dangerous behaviours in the pursuit of belonging, such as involvement in street gangs or organized crime groups (Todd & Young, 1994). Williams (2009) similarly finds that ostracism, and group-based ostracism likewise leads to resignation, alienation, helplessness, and depression if the individual is not able to adapt to the detection of ostracism. Ostracism is an interpersonally

aversive behavior because it threatens four of our fundamental needs: the need to belong, the need to maintain a certain level of self-esteem, to perceive a degree of control of one's environment, and to feel recognized and worthy of attention (K. D. Williams, 2009).

Indeed, the benefits of increased social relationships for people with life-threatening illnesses include lower cognitive decline with aging, greater resistance to infectious disease, and better prognoses when managing life-threatening illnesses (S. Cohen & Janicki-Deverts, 2009). Cohen and Janicki-Deverts (2009) hypothesize that the association of social integration and health may be mediated by perceived social support. In another study, participants were randomly assigned to receive messages of social exclusion, researchers documented significant and large decreases in intelligence measures (IQ, GRE), with evidence of a reduction in both speed and accuracy of cognitive abilities (Baumeister et al., 2002). These effects were not mediated by mood, and were only apparent when participants received messages of social exclusion, but not nonsocial misfortune, like accidents or injuries.

Measurement of Belonging

Some psychometric evidence has already been collected concerning the construct of social belonging. For example, Newcomb (1990) finds evidence of "connectedness" as a higher order factor of social support and loneliness, where he views these two factors as opposite ends of the psychosocial construct of "connectedness." According to their work, the unified concept is moderately stable, modestly transactional, and is generalizable across interpersonal situation. Related, extensive research confirms the

importance of peer and adult connections for youth in British Columbia as a way to support student well-being (Gadermann et al., 2016; Guhn et al., 2012; Oberle, Schonert-Reichl, Guhn, Zumbo, & Hertzman, 2014). Expanding to the field of self-psychology, Lee and Robbins (1998) developed the "Social Connectedness Scale" (SCS) to measure interpersonal closeness between a person and their social world of friends, peers, and society and the challenge of maintaining a sense of closeness. Out of concern of the growing feelings of isolation and loneliness experienced by adolescents and immigrants, they suggest that connectedness is made of three components: companionship, affiliation, and connectedness distributed across a two-factor model (Lee & Robbins, 1995), which is consistent with the prior work of Patton et. al (1982). Their findings also align to other theoretical research on connectedness and development literature on belongingness and attachment (Cherry, 1994) and suggest that social connectedness is relevant to lower trait anxiety; accounting for roughly 16% of the total variance in trait anxiety. Lee and Robbins (1995) work suggests that social belongingness is indeed related to other constructs of attachment, loneliness, or perceived social support, but that it is still distinct. Using a series of factor analyses to test the internal structure of social belonging, they find a two- factor model of Social Assurance and Social Connectedness (Cronbach's α =91 and α =.77 respectively). The test-retest correlations were estimated with separate samples during 2-week intervals and were .98 and .84 respectively with sufficient goodness of fit statistics. Relatedly, the construct of "loneliness" is considered to have a three-factor structure consistent with Brewer and Gardner's three-part construct of self: intimate self, relational self, and

collective self (Cacioppo & Patrick, 2009), with all three aspects being highly correlated with one another. Loneliness is also considered a distinct construct from depression, although they are highly related. Cacioppo and Patrick (2009) describe loneliness and depression as almost opposites where loneliness triggers one to do something to get out of discomfort, like hunger, whereas depression makes one apathetic and holds one back.

Indeed Baumeister and Leary (1995) also suggest that social exclusion may be one of the largest factors of anxiety. They suggest that social isolation may be the only objective factor correlated with subjective well-being (1992). Evidence from the creation of the Social Connectedness Scale shows the scale to be correlated to the Collective Self-Esteem Scale (.55*), the Social Support Questionnaire – Network (.31*), Social Support Questionnaire – Satisfaction (.34*), and the State-Trait Anxiety Inventory – Trait Scale (-.63*). Williams (2009) also finds that the construct of belonging is tied up with self-esteem and that the two might be considered as an "inclusionary need" cluster.

Others have also taken up the question of belonging in the context of children and schools. Goodenow (1993) developed the "Psychological Sense of School Membership" (PSSM) scale, which is an 18-item scale that is correlated with self-report measures of motivation, grades, and teacher-rated effort levels. The PSSM scale builds upon Weiner's (1990) theory that belonging is key to motivation. Likewise in the school context, belonging has been well documented to be related to school retention and participation of "at-risk" students (Finn, 1989). Evidence suggests that those who do not "fit in" or who see themselves outside of mainstream culture are more likely to

disengage and drop-out (Wehlage, Rutter, & Smith, 1989). Interestingly, Goodenow's research also suggests that feelings of belonging may be related to ethnic representation within a school. In those schools studied, when minority students were the majority within a school, minority students had a much higher sense of social belonging. Likewise, in schools where representation was roughly proportional, there was not clear evidence of group differences in belonging.

Moreover empirical evidence speaks to higher incidences of psychopathology for children who experience rejection (Hamachek, 1987). Meanwhile people who experience social isolation are similarly likely to be at risk for illness or early death as those with high blood pressure, obesity, lack of exercise, or smoking. However there is not a clear causal path for how these negative effects happen, but rather it is regarded as more of a grinding down, wear-and-tear type of process (Cacioppo & Patrick, 2009). Indeed, in a study of older adults, it is the sense of loneliness, not a lack of social support, that uniquely predicts depressive symptomatology and chronic health problems, possibly as a result of "self-protective" behaviors becoming ingrained and a constant source of stress with increased production of epinephrine, which is a stress hormone that can make one susceptible to illness (Cacioppo & Patrick, 2009).

Using a shortened version of the scale developed by Walton and Cohen (2007), Cohen, Garcia, and Apfel (2009) designed a structured writing assignment to investigate students' self-affirming value for younger students. This revised scale had two subscales; self-perceived social belonging (5 items; $\alpha = 0.77$) and self-perceived ability to succeed (4 items; $\alpha = 0.66$). Each item was on a 6-point scale from 1 (strongly

disagree) to 6 (strongly agree). The overall Cronbach's alpha was reported at .79. The authors assessed the two subscales using a principal components analysis and varimax rotation on all 9 items, with results indicating two distinct scales (social belonging: eigenvalue 3.52; ability to succeed: eigenvalue 1.23) with loadings ranging from .56 to .77.

Hoffman et al. (2002) developed the "Sense of Belonging" scale in part related to concern of college attrition, with the understanding that if students develop a sense of belonging at school they are more likely to complete their studies. Hoffman et al. (2002) adopt the definition of social belonging put forward by Hagerty et al. as "the experience of personal involvement in a system or environment so that persons feel themselves to be an integral part of that system or environment" (1992, p. 173). Similarly, sense of belonging is considered to be an aspect of interpersonal relatedness that closely associated with social support and dissimilar to loneliness (Hagerty et al., 1992). Their original instrument included 85 items (50 student/peer relationships; 35 peer/faculty relationships) and was conducted with first-year university students in Rhode Island, USA. Their analysis sought to test the effect of participation in certain "learning communities", but also included a factor analysis to refine their scale. They found four underlying dimensions: perceived classroom comfort, perceived isolation, perceived academic support, and perceived social support and results explained 68.5% of the variance. The peer/faculty items had three dimensions: empathetic understanding, perceived faculty academic support/comfort, and perceived faculty social support/comfort which explained 73.3% of the variance. After examining the loadings,

the scale was then reduced to 26 items with five dimensions: Perceived Social Support (α = .87), Perceived Faculty Support/Comfort (α = .87), Perceived Classroom Comfort (α = .90); Perceived Isolation (α = .82), and Empathetic Faculty (α = .85) and explained 63.3% of the variance.

Expanding on Hoffman et al.'s work, Tovar and Simon (2010) investigated measurement invariance of the sense of belonging scale across class standing in university under the belief that students' perception of belonging would change over time. Tovar and Simon, recognizing that Hoffman et al.'s work only used principal components analysis, and so they also conducted Confirmatory and Exploratory Factor Analyses (EFA/CFA) in order to have more thorough psychometric evidence for the SBS. Using more appropriate factor analyses tests, and examining factor fit, Tovar and Simon, proposes a revised Sense of Belonging Scale resulting in a three-factor, 16 item instrument (α = .90). These are: perceived faculty understanding/comfort (7 items; α = .89); Perceived Peer Support (6 items; α = .84); and Perceived Classroom Comfort (3 items; α = .93). Interestingly, Tovar and Simon did not find any statistically significant difference between students' response to the SBS scale as a result of class standing (recent entry into university vs. more senior students), which suggests that the SBS may be invariant based upon time at university and it was determined to have the same latent mean score of sense of belonging.

Using the Sense of Belonging Scale, some research has been conducted to examine how racial climate affects students' sense of belonging (Johnson et al., 2007; Nuñez, 2009), finding that hostile racial climates cause a decreased sense of belonging,

but that early, positive transition experiences can enhance a sense of belonging. Nuñez (2009) also analyzed sense of belonging with Latino students in the US and found the models to be predictive, with second-generation immigration status being predictive of a decreased sense of belonging in comparison to either first-generation immigrants or third-generation.

PISA's Measure of Social Belonging in School

In addition to the academic and other sociological components, PISA also includes a 6-item scale assessing students' sense of belonging in school (see Table 1-3.1 for items). PISA began measuring sense of belonging in 2000 out of an effort to understand the connection between student engagement (belonging and participation) on academic outcomes. For adolescents, a sense of belonging in school is related to academic success in terms of achievement, attendance/truancy, school completion, feelings of security, and psychological and sociological development (Goodenow, 1993; Jethwani-Keyser, 2008). Likewise, especially for immigrants, schools play a key role in the process of cultural adaptation as schools are often the first social and cultural institutions that children have contact with in their new country (Chiu, Pong, Mori, & Chow, 2012). However, for all students, schools are the place where students spend the majority of their waking hours. As such, I take for granted that a students' sense of belonging in school is in many ways a proxy for their overall sense of belonging. I do not attempt to parse out different "types" of belonging. Yet, I recognize that "belonging" as a general concept includes more than just school connection and encompasses peers, families, and other group identities. So while I frequently use the

shorthand "social belonging", it is important to be clear that the construct explicitly captures sense of belonging in school.

Published PISA results regarding sense of belonging at school analyze family, school, and student variables and find that most of the variance is attributable to the individual level (~90%), ~3% at the school level, and ~8% at the country level (OECD, 2017b). Descriptive results based on migration background, find those with a more recent migration history tend to have lower rates of belonging as do those with lower socio-economic status (SES). Published OECD reports also suggest that when teachers are perceived to support them fairly and support their learning, students have an increased sense of belonging. In general, more socio-economically advantaged students reported higher levels of belonging, but results were mixed for student sex: Australia, Denmark, Finland, Ireland, Norway, the UK, and the US all had higher rates for boys while Jordan, Qatar, and Turkey had higher rates for girls. There is not detailed information publicly available about the development of the PISA measure of belonging, which warrants further study.

Research conducted in many countries and schools requires representative sampling. In administering the PISA, the OECD uses a two-stage, stratified, sampling framework (and a three-stage framework for Russia, who is not discussed in this dissertation). The first-stage consists of all eligible schools from the national sampling frame with probabilities proportional to a measure of size; which itself is a function of the estimated number of eligible students in the school. PISA uses 150 schools that are representative based on SES and student intake. PISA uses a number of explicit strata

categories that differ based on country, but include such things as region, racial demographics (ensuring representation of certain groups), language groups, urbanicity, and others (see PISA technical report for complete list by country: OECD, 2017b). Stage-two then consists of selecting students within the schools using a target cluster size specification of at least 20, but generally 35 for paper-tests and 42 for online tests. It is relevant to note OECD excludes students with mental or functional ability issues assessed by local professional staff, and those students who do not speak the test language. Less than 2% of the total sample were excluded (OECD, 2017b). School response rates of at least 85% were required, and a minimum of an 80% completion rate for students within schools, however there were few issues with non-respondents. Using sample weights, the samples are accordingly adjusted to provide a representative samples of 15-year old students in each country; where weights are reciprocal to the inverse of their probability of sampling. In the case of theoretical constructs, like sense of belonging, cross-country measurement invariance is assessed using Rasch modeling on the overall sample only.

Integration and the Policy Dimension of Social Belonging

This work is explicitly and implicitly informed by the field of political philosophy and political science, and I often adopt the frame of *political liberalism* (not to be confused with "liberal" parties or economic liberalism) consistent with the likes of John Rawls, Charles Taylor, and others. Through the act of measurement itself, by seeking to measure an inherently political concept like social belonging, and by

conducting analysis of measurement invariance based on group belonging, I am necessarily engaging in debates about political philosophy and the act of policy formation. Indeed, the act of measuring itself already reflects a certain set of values, including political values. Consistent with Rawls, political philosophy ought to be practical, help citizens orient themselves in the social world, probe the limits of practicable political possibility, and provide reconciliation. To paraphrase Rosseau, philosophy imagines how laws might be. By seeking to understand the context of social belonging and its connection to migration background, I am explicitly making certain policy-type assumptions. Namely, I assume that we want all denizens to have a sense of belonging. I assume that public policy is one such political lever that acts upon denizens to shape one's social world. And I assume that by deepening our understanding of a social phenomenon, in this case social belonging and migration background, we might be able to improve our political world to maximize how laws might be.

While not always the case, immigrants, as members of a non-dominant group, often face similar individual and institutional discriminations as "native" non-dominant groups. In many cases, this means that immigrants face the additional burden of "racism" in their host countries, which result in adverse health outcomes and restricted socioeconomic opportunities and mobility. In many cases immigrants have entered new host societies under the auspices of cheap, and often assumed temporary, labour that results in a painful process of acquiring new spatial and social senses of belonging. Meanwhile the host-society likewise experiences a cultural shift with the introduction of new immigrant customs (Sandercock, 2003). A particularly interesting finding is that

immigrants, even low socio-economic status (SES) immigrants, often have better health status than native-born populations, but their health status deteriorates with increased length of stay (E. J. Williams, 1995). Williams (1995) suggests this may be because racial groups often captures differences in power, status, and resources in the US, and racial segregation and accordingly racial differences in SES are not just a confounder of health, but are in fact a part of the causal pathway with race as an antecedent and a determinant of SES. Similarly, Hummer et al. (1999) find that native-born blacks have the highest odds of death, while foreign-born blacks and Asians have lower death odds for older adults, with Mexican Americans and other Hispanics with intermediate risks in the USA. Their findings suggest that mortality is influenced by nativity with foreign-born individuals experiencing lower risks, and similar to Williams (1995), that their acculturation to the US wears away some of their advantage. In the domain of health research, that means that nativity ought to be considered as a control factor.

For immigrants, the idea of "belonging" has is studied differently than what I have previously discussed. Instead, immigrant belonging is discussed in the context of "integration" and generally concerns policy, economic, socio-political, and psychological questions. Then, from this framework, the debate around multiculturalism in a theoretical sense becomes tied to that of immigrant belonging. Historically, this frame is connected to the shift towards "civic" nations; rather than "ethnic" nations, where race or colour are not assumed, at least theoretically, to be barriers of integration. However, this is not to say that civic nations are free from a cultural identity. Claims for rights by immigrants do not have to be considered a

rejection of integration, but can instead be a kind of renegotiation of the terms (Kymlicka, 1996). Respect for minority rights is actually an enlargement of freedoms, since personal freedoms can, and often are, tied up with claims for community and/or cultural respect. Cultures have value for their inherent existence, and for their access to meaningful ways of living and knowing that often depend upon access and understanding to cultures (Dworkin, 1985).

Much scholarship of immigrant social cohesion and belonging was done with specific interests, such as interethnic marriage, religiosity, language proficiency, residential segregation, ethnic economic inequalities, and ethnic identity, however there has been a lack of study around the cultural patterns and values between and among immigrants and their new societies (Nee & Alba, 2003; van Tubergen, 2006). To be clear, I refer to the process of immigrants adapting into their mainstream host society when I use the term "integration." Civic integration assumes certain principles: First, that employment/economic subsistence is important and second, that a respect for liberal democratic values from migrants and the host society is present. These values include liberty, democracy, equality, and the rule of law. Third, at least an elementary proficiency of the host societies' language, history, and institutions are key, and fourth, anti-discrimination policies are necessary for increased integration (Council of the European Union, 2004; Joppke, 2007). Thus, a cohesive society embodies a common sense of belonging and identity with some level of trust.

Currently many of the world's immigrant-destination countries are concerned about migration and its effects on social cohesion. In fact, Amy Gutmann says, "it is

hard to find a democratic or democratizing society these days that is not the site of some significant controversy over whether and how its public institutions should better recognize the identities of cultural and disadvantaged minorities" (Taylor et al., 1994, p. 3). In Germany, for example, data indicates that immigration may be the biggest field of contention, ahead of employment/labour issues, war/peace, democracy, or environmental issues (Rucht & Neidhardt, 2001). Wright and Bloemraad (2012) conduct a variety of cross-national studies around social inclusion, political inclusion, and political engagement and present evidence that multicultural policy frameworks in most cases foster engagement between immigrants and society and their host government, and in no cases hinders such engagement. Their findings reject the dichotomy of integration and national attachment on one hand and multiculturalism on the other, or that immigrants and minorities somehow live "parallel lives" (e.g. the Cantor report in the UK or the notion of Parallelgesellschaften in Germany).

The debates around migration flows thus represents more than just a flashpoint of opposing interests in society, but also of the (re)conceptualization of national identities. Similarly, it is not just migrants who are managed in societies, but also the nativist response, particularly of right-wing extremist and whether or not there is a politicization of cleavages and what kinds of political space are afforded to them as political actors (Koopmans, Statham, Giugni, & Passy, 2005). Indeed, evidence suggests that members of minority communities have an increased feeling of belonging and engagement with the polity when their respective cultures are accommodated, since successful immigration is not only a matter of the individual choices migrants make to

join the new polity, but also about the reception they receive (Bloemraad, 2006). By examining policy regimes, Wright and Bloemraad (2012) go beyond the individual-level determinants of affect and the macro-structures of political party, which offers an expanded approach to analyze policy levers on individual's lives. Since immigrants as minorities are by definition in a position of cultural inequality, the extension of policy regimes that offer citizenship or residency are not enough, but instead inclusion is only to be achieved through recognition and accommodating cultural beliefs and practices. In this case, equality in opportunity is not concerned with identical treatment, but rather differential treatment to ensure equal access to jobs, services, and civic participation (Parekh, 2006). Indeed, for migrants, "self-segregation" within host-societies sometimes has little to do preferences, and more to do with discrimination in housing or access to opportunities that prevent migrants from being able to "freely" make choices like their native-born counterparts.

By encouraging membership through norms and policies, immigrants may adopt nested or hyphenated identities that attenuate the political problems of pluralism. Indeed, Kesler and Bloemraad (2010) find that people are actually less likely to report declines in trust, organizational membership, and political participation in multicultural societies compared to non-multicultural societies. This finding is echoed by Bauböck (2003) who argues that shared identities in societies are necessary, but that citizenship should allow for overlapping identities. Using Bauböck's (2003) framework, we ought to conceive of a catalyst model where immigration is a catalyst that sets off a broad process of self-transformation towards a more pluralistic and cosmopolitan identity of

the whole of society. The difficulty is how to foster a mass restructuring of identity beyond the migration of a few "elites." Wright and Bloemraad (2012) overall find that all immigrants, even those who have not or are not seeking citizenship, feel more welcomed in societies with inclusive citizenship laws due to their symbolic legitimacy. Wright and Bloemraad conclude that first, the claim that multicultural policies fail to promote civic integration is false and second, that the repeal of multicultural policies will further marginalize immigrant populations to the detriments of civic inclusion and political legitimacy. Indeed, they find a correlation of .70 ($p<.01$) between the presence of multicultural policies and the increase in citizenship (Bloemraad & Wright, 2014).

One of the largest problems from a policy perspective is that most countries have managed their influx of immigrants on an ad-hoc basis using institutions created for other purposes, resulting in incoherent sub-systems (Freeman, 2004). Yet, in recent years, multiculturalism has been chastised and has been blamed for residential ghettoization, social isolation of migrants, poor economic integration, poor educational outcomes, welfare dependency, illiberal practices by migrants, inequality – particularly for women, and political/religious radicalism (Banting & Kymlicka, 2013). The assumption being that if immigrants have easy access to equal rights, then they have no incentive to make an effort at integration, which then leads to social and economic marginalization (Koopmans, 2010). This has resulted in a backlash to multiculturalism in countries like the Netherlands, where countries are caught in populist waves and are concerned about being left behind by globalization (Entzinger, 2014). Immigrants do not demand the same rights or autonomy as national or indigenous minorities. Simply,

citizens in multicultural democracies may have a multitude of self-governing religious, ethnic, or linguistics polities (Baubock, 2003).

One of the largest critics of multiculturalism has been Putnam, who argues that in the short-term increased immigration and ethnic diversity is likely to reduce social solidarity and social capital as a result of diverse neighborhoods "hunkering down", but that in the long-term societies will likely garner benefits of diversity (2007). Putnam argues that the phenomena of "hunkering down" results in overall decreased levels of trust and sometimes decreased investment in public goods for more diverse neighborhoods, not only between immigrants and native-born persons, but also among native-born persons. This same sentiment is echoed by Coleman (1988) who laments the ever-growing atomization of society and community closure and by the Council of Europe who declared "what had until recently been a preferred policy approach, conveyed in shorthand as 'multiculturalism', has been found inadequate" (Council of Europe, 2008, p. 9). Even though immigrants often bring enormous benefits, for example immigrants received three to four times America's Nobel Laureates, National Academy of Science members, Academy Award film directors and winners of Kennedy Center awards compared to native-born persons in the US. Thus, Putnam argues, the main challenge of increasingly diverse societies is to create a broader sense of who "we" is.

There has been considerable refute of Coleman and Putnam's works. Using the World Values Survey and the Danish Social Capital Survey, Bjørnskov (2007) finds ethnic heterogeneity is not significantly related to a respondent's expression of trust.

This finding is further confirmed using the Latinobarometer and Afrobarometer surveys (Bjørnskov, 2008). Bjørnskov finds trust to be a stable measure although they note the importance of taking endogeneity seriously (2007). Their work, also includes interesting societal level effects on trust, such as Protestantism or the presence of a monarchy leading to increased trust, while post-communist societies have reduced levels of trust. Also, of particular interest is their finding that the transmission of trust may have an element of cultural inheritance. Hooghe et al. (2009) use another measure of immigrant diversity and similarly find ethnic diversity is unrelated to generalized trust using multilevel modelling. Indeed Portes and Vickstrom (2011) also argue that the notion of "communitarianism" as defined by Putnam and others is not ideal for smooth operation of policy in our present world, and rather that communitarianism is merely a byproduct of more basic structural factors of which racial homogeneity, education, and economic equality are critical. One might use the distinction of "mechanical solidarity" compared to "organic solidarity."

Indeed, recent evidence by Bloemraad (2015) suggests that countries with multicultural policies and stronger pluralism are the same places where immigrants are more likely to become citizens, have trust in political institutions, and feel a sense of attachment to the national identity. Likewise, their research finds little evidence of majority backlashes against multicultural policies, no positive or negative effect for second generation immigrants, some evidence for convergence with third generations or later, inconclusive evidence in terms of socioeconomic integration, and only some evidence that such policies expand the notion of inclusive membership (Bloemraad,

2015; Bloemraad & Wright, 2014). Related, Wright (2011) finds that increased amounts of spending on social welfare actually retards any potential economic effects posed by an influx of migrants, and that these increased levels of spending are related to more inclusive definitions of national communities. Bloemraad (2015) uses her findings to argue that the participatory aspect of citizenship needs to be central to future research in social theory. In addition to the useful review of multicultural policies, Bloemraad's work makes clear that belonging, status, rights, and participation are not necessarily related to one another, and that greater nuance is needed, particularly around the normative claims embedded in citizenship; contrary to the work of Joppke (2010) who suggests that citizenship is first about state membership in a political body with citizenship only secondarily being about equality. Rather Bloemraad, Korteweg, and Yurakul (2008) disaggregate citizenship into four dimensions: legal status, political engagement, rights, and a sense of belonging. Ultimately the work by Bloemraad argues that multiculturalism is important not only to accommodate and produce nested identities attached to the state, but are also an important philosophical critique to the rights-based liberalism framework. The communitarian alternative to liberalism states that people achieve meaning and identity from their particular society and cultural communities, and individual agency is meaningless without recognition of our communities, and it is this link that ties multiculturalism directly to the politics of citizenship and belonging (Parekh, 2006; Taylor et al., 1994). But what is perhaps unique in the modern context, is not the implicit need for recognition, but a creation of conditions in which the attempt to be recognized can fail (Taylor et al., 1994).

Koopmans et al. (2005) conceive of citizenship along two continuums: on the vertical continuum is the idea of citizenship as a form of ethnic bonds compared to that which emphasizes rights and citizenship on a territorial basis, and, on the horizontal scale, the conception of citizenship as conformity to a single culture shared by all citizens compared to a pluralist conception where each person retains their diversity and possess a multitude of expression. While the extremes are hardly achieved along either continuum, they are perhaps useful conceptions about the ways in which immigrant belonging becomes framed.

Contextually, many countries are troubled by a sense of postcolonial guilt and constant fear of being accused of racism. As such, on the one hand policy makers may view migrants as incapable of ameliorating their own problems and are in need of assistance and on the other hand, a wariness to use the state's power to push migrants to alleviate their disadvantage given many countries' troubled histories. This has created a cross-national divergence of structures of citizenship and identity that is also influenced by the demand for the right to difference experienced within a country, the visibility of protest, and general public legitimacy. In Koopmans et al. (2005) research, they see a convergence of countries shifting towards civic, territorial conceptions of citizenship and a reversal of differentiated citizenship and assimilationist approaches to allow for stronger cultural rights and differences. One of the more surprising findings from their research is that strong transnational orientations are associated with citizenship regimes that put in place strong barriers for migrants, rather than being a result of open, pluralistic receiving countries. It is important to keep in mind though,

that the conflicts over culture and belonging are not just about values, but are also linked to material aspects of society. The demands made by and made of migrants in public education, welfare, and other services makes these tensions much more salient.

Interestingly, Kesler and Bloemraad (2010) find that multicultural policies and greater economic equality both appear to reverse the loss of social capital caused by increases in diversity talked about by Putnam and others. The theory is that social cohesion is tied up with social capital, which analyzes relationships between people of the same groups ("bonding") and with other groups ("bridging"), and that strong bonding within multicultural communities offers a platform to build bridges with additional communities (Spencer & Cooper, 2006). In fact, a review done by Berry (2005) and a separate meta-analysis of 52 psychological studies by Nguyen and Benet-Martínez (2013) both find that people that fuse their heritage culture with a new attachment to a national society have higher levels of tolerance and self-reported levels of well-being than others with unitary attachments. Indeed, bi-cultural peoples often have a stronger link in the integration procession than those with only a single culture (dominant or heritage), and that acculturation is facilitated and psychological and emotional well-being are supported by encouraging people to use their tiered identities to navigate new and old cultures (Nguyen & Benet-Martínez, 2013). Bi-cultural persons have strong support networks (Mok, Morris, Benet-Martínez, & Karakitapoğlu-Aygün, 2007), and the experience of negotiating multiple cultures supports intellectual flexibility and creativity (Benet-Martínez, Lee, & Leu, 2006; Tadmor, Tetlock, & Peng, 2009). Likewise, the encouragement of multi-tiered identities has enormous benefits for

students in particular. Nee and Alba (2003) find that students being able to use their own native slang, such as Moroccan Arabic or Tamazight in a school environment added to their sense of belonging. This type of belonging is critical to help students realize their full potential and lay claim to their entitled full membership in a new society. The link between socioeconomic origins and educational outcomes varies by national contexts (Breen & Jonsson, 2005), which suggests that comparative methodologies to examine immigrants in schools can help uncover how to reduce discrepancies in opportunities. Indeed, Australia and Canada stand out in an international context given the parity of educational outcomes between native born and immigrant origin students, which is often attributed to differences in their selective immigration regimes (Nee & Alba, 2003).

Methodologically, many scholars take a case-study approach, generally seeking to understand immigration of a particular national group into a particular country. However, this often results in many hidden national assumptions that can be taken for granted, as such Alba and Foner (2014) suggest that a comparative approach can shed light on these "invisible constants." Comparative analyses thus help to expose a nation's pre-existing idea of self-understanding and frame policy of inclusion around institutional arrangements and social interactions (Bertossi & Duyvendak, 2012). For example, one larger frame often used is the "settler/non-settler" society dichotomy to compare countries such as Canada or the US to European countries, with the belief that "settler" societies are quicker to extend the umbrella of belonging to other groups. But of course, the central debate is not whether or not national ideologies or models exist,

they do. Rather, the question is to what extent these frames influence the outcomes for immigrant minorities and their overall belonging. With an increased acceptance that policy regimes do matter for social belonging, we thus must answer how governments can foster these policies (Harell & Stolle, 2010; Wright, 2011). Instead, social scientists must peel back the layers of what it means to be a "migrant," since the legal origin of a migrant results in substantially different outcomes and opportunities, which recognizes that the rights of permanent residence, temporary visa holders, refugees, skilled migrants, and undocumented migrants are often quite variable.

In terms of immigrant integration, one of the most common measures used is that of "trust" and especially political trust. The theory being, that analysis of political trust will help uncover to what extent migrants experience alienation or attachment to mainstream institutions, which are central to integration (Joppke, 2007; Maxwell, 2010). Maxwell (2010) finds that first-generation immigrants often have the highest level of political trust, presumably because they undergo a disruptive change, have lower expectations and accordingly have more positive evaluations, meanwhile native-born and second-generation migrants, who are raised in the same society, have similar levels of political trust. However, second-generation migrants face their own set of challenges, particularly around stigmatization and discrimination, with feelings of cultural entrapment as neither fully a part of their parent's country of origin nor their host country. This results in increased feelings around discrimination (Maxwell, 2010). Indeed, while unsurprisingly, migrants face initial challenges of integrating, yet over time their outcomes converge with native-born persons, with some of the key

determinants being acquisition of citizenship and adoption of host society customs and especially language(s) since language is a key determinant of social capital (Maxwell, 2010; van Tubergen, 2006). van Tubergen (2006) finds that when host societies are more discriminatory towards migrants, this in turn reduces the likelihood of migrants learning the host language. Similarly, Brubaker (1998) finds that in countries where immigrants have easier access to citizenship, they are more likely to participate in politics and advance their interests in mainstream society.

Immigration is now a permanent feature of our societies (Council of the European Union, 2004). Thus, it behooves us to foster a well-managed process of integration that successfully supports migrants at the individual, family, community, and state levels, which all simultaneously occur. By fostering social belonging among migrants and native-born persons, we can collectively reap the benefits of increased cohesion, stronger economies, feelings of security, and cultural diversity. It is true that in some ways "full belonging" is more of an ideal than it might be an actual reality. Instead, it may be better to think of societal identity as an "always contested engagement with and redefined notion" with a shared destiny of communities (Sandercock, 2003, p. 151). However, it is important to engage seriously with the politics of recognition and the importance of managing difference in a way that is consistent with its roots of universal dignity (Taylor et al., 1994). Indeed, as Charles Taylor (1993) suggests, we need a theory of "deep diversity" that can accommodate not only multiple nations, and distinct cultural groups, but also diverse ways of living and the ways in which members can belong to the large polity. For Parekh (2006), the way forward is to

accept "the full force of moral and cultural pluralism and acknowledge that the good life can be lived in several different ways, some better than others in certain respects but none is the best." (pg. 110). Perhaps, with respect to social belonging, the question of "who are we?" should be abandoned altogether for the more radical question of "how are we all going to live together?"

Chapter 3: Migration Background in PISA's Measure of Social Belonging: Using A Diffractive Lens to Interpret Multi-Method DIF Studies

Introduction

The purposes of this paper are twofold. The first is to investigate measurement invariance in an international context using the Programme for International Student Assessment's (PISA's) measure of social belonging as it relates to migration background, which has yet to be explored. The second is to introduce a diffractive lens to interpret a multi-method investigation of Differential Item Functioning (DIF). In this case, the multi-method strategy involved two statistical methods: the alignment method in conjunction with ordinal logistic regression (OLR) in the case of multiple group comparisons. The aim is not to recommend one DIF technique over the other, but rather to demonstrate how the use of a diffractive lens on measurement invariance can reveal insights about how individuals interact with survey items based upon their social context in an international context.

To accomplish these goals, I structure the paper in the following manner. I first explain the diffractive lens that shapes the framework and paradigm of the DIF analysis, then I introduce the construct of social belonging and why it is worth exploring while making explicit connections to the way in which migration background is relevant to this discussion. Next, the alignment method and OLR are briefly described. The alignment method is a particularly useful tool in the comparison of migration background in an international context because of the many groups involved.

Indeed, if this analysis was concerned with only a single-country the alignment method would be unnecessary. OLR allows a more nuanced investigation of the items of the scale. Next, the findings of a study of measurement invariance of PISA's social belonging measure as related to migration background for English speaking test-takers are reported. Finally, in the Discussion section I situate the findings in the broader context of measurement invariance related to migration background, multi-method DIF approaches along with the use of the diffractive methods, and I conclude with a remarks about both the policy and measurement implications of this research and situate these findings within an ecological model of item response (Zumbo et al., 2015).

Diffractive Methodology

The use of the alignment method and logistic regression DIF is intended as a kind of "diffractive methodology." To draw upon a physics example, one might think about light as a construct, being investigated within two different mediums (i.e., two different methodologies): air and water. The sunlight itself is the same substance, but it behaves or manifests itself uniquely within air or water. Likewise, when the two mediums are placed side-by-side shining a single light, one can see a distinct bend or diffraction occur. Neither medium can fully capture the "truth" of how light behaves in its entirety, but rather it is the contradictions and tensions between the two approaches that more fully captures the construct of interest. In the same way, I seek to investigate DIF of social belonging based upon migration background (the construct) using two different methodologies: the alignment method and OLR. Given that researchers have already established that item- and test-level analyses do not necessarily have the same

conclusions about construct or measurement comparability (Oliveri & Ercikan, 2011; Zumbo, 2003), it is useful to consider not only what conclusions are shared across methodologies, but how their differences shed light on individuals and/or groups and their interactions within a measure of belonging (the scale: via alignment), and the components of the measure (the items: via OLR). Building on the work of Barad (2007), Dixon-Román (2017) develops the philosophical justification for diffraction within social sciences, which is interested as much in the produced differences of methodological approaches as that which is shared. To be clear, diffraction does not mean "mixed-methods", which focus on corroboration of findings, but instead diffraction focuses on the tensions and contradictions between data and findings when viewed from multiple methodologies. I use this lens to guide the interpretation to the results of both the alignment and OLR.

Social Belonging

Zumbo et al. (2015) argue that one needs to examine explanatory factors related to item-responses, including social context. I argue that immigration background is one such factor that is important when considering social belonging and how an individual interacts with a survey/items about belonging. This framework expands the research of social belonging beyond statistical comparability to examine the "how" and "why" of social belonging with respect to immigration background as a part of the validation process.

According to the belongingness hypothesis, humans have a minimum need for lasting, positive, and significant interpersonal relationships dependent upon two criteria: a need for frequent and affectively positive interactions with at least a few people; and second, these interactions should occur in a stable and lasting framework of concern for one another's welfare (Baumeister & Leary, 1995). Thus, belonging is greater than mere affiliation or social attachment, and is of critical importance for one's socio-emotional well-being. Social belonging, defined as "a sense of having a positive relationship with others" (Baumeister & Leary, 1995) is a central human need, and the lack of belonging, or exclusion, harms people in a number of ways including, subjective well-being (Lyubomirsky et al., 2005), intellectual achievement (Walton & Cohen, 2007), health and immune functioning (Berkman & Syme, 1979), and self-control (Baumeister et al., 2002).

While the need to belong seems relatively universal, the difficulties of achieving a sense of belonging is unevenly felt. Indeed awareness of one's belonging to a stigmatized group, such as race, religion, gender, ethnicity, or sexual orientation, in a particular value domain often results in decreased outcomes, such as academic performance (Steele, 1997; Steele & Aronson, 1995). Similarly, for persons a part of stigmatized groups, the question of personal belonging is part and parcel with the question of group belonging at large (G. L. Cohen & Garcia, 2005). For already marginalized and stigmatized individuals, the question of acceptance and belonging is considered the "central feature" of one's life, since it often defines and shapes one's entire life trajectory (Goffman, 1986). Immigrants, as members of a non-dominant

group, often face similar individual and institutional discriminations as "native" non-dominant groups. In many cases, this means that immigrants face the additional burden of prejudice in their host countries, which result in adverse health outcomes and restricted socioeconomic opportunities and mobility.

The idea of "belonging" for immigrants has not been studied in the same ways in which it has for other non-dominant groups. Instead, immigrant belonging is often tied up within the conversation of "integration" and is most often treated as an economic-policy question rather than one of social well-being. Most analysis of immigrant social cohesion and belonging has been taken up with specific interest in interethnic marriage, religiosity, language proficiency, residential segregation, ethnic economic inequalities, and ethnic identity, but with little attention to the cultural patterns and values between and among immigrants and the host society (Nee & Alba, 2003; van Tubergen, 2006). As a policy question, and given the nature of immigrants as necessarily "foreign" in some capacity, immigrant belonging becomes a particularly contentious example of how we foster and encourage belonging in our society. As Amy Gutmann states, "it is hard to find a democratic or democratizing society these days that is not the site of some significant controversy over whether and how its public institutions should better recognize the identities of cultural and disadvantaged minorities" (Taylor et al., 1994, p. 3).

Assessments, like PISA, are frequently used in countries to inform policy decisions, such as resource allocation, or educational practice, such as curriculum planning (Howie & Plomp, 2006). However, careful attention should be paid to cross-

cultural comparisons. Existing research confirms the presence of DIF in PISA in Mathematics (Yildirim & Berberoğlu, 2009), Science (Kankaraš & Moors, 2014), in a cultural context, but that DIF can be mitigated. Similarly, DIF has been found across immigrant groups on Reading (Da Costa & Araújo, 2012), which reaffirms the need to think carefully of the cultural setting of testing. In addition to the academic and demographic components, PISA includes sociological scales; including belonging. These questions are asked in schools because it is one of the most important settings that shape child development, health, and well-being (Eccles & Roeser, 2011). Indeed, students in Organization for Economic Cooperation and Development (OECD) countries such as Canada or the US, spend approximately 7,500 hours on average in a classroom (OECD, 2017a) with their interactions affected by connections with peers, teachers, staff, and the broader school norms and values that construct climate (Blum & Libbey, 2004). A positive environment that supports social belonging contributes to healthy students, and a stressful and alienated experience can cause mental health problems and failure (Hamre & Pianta, 2005).

Measurement Invariance

To investigate measurement invariance based upon immigration background and the way it interacts with PISA items, I use the "alignment method" and OLR to investigate the presence of DIF, which is when "some characteristic of the test item and/or testing situation that is not relevant to the underlying ability of interest and hence test purpose" (Zumbo et al., 2015, p. 137) occurs, and hence the item (or scale) may be measuring different abilities across sub-groups. The presence of DIF, suggests

that items do not adequately capture the construct under investigation equally across groups, and compromises one's ability to draw conclusions. As such, inferences cannot be attributed to just a participants performance, but may be due to another measurement artifact (F. Chen & West, 2008). More broadly, DIF is central to the practice of validation itself (Gómez-Benito et al., 2018) since it compromises our ability to interpret a test score at all, and accordingly draw inferences.

Measurement invariance is a requirement in order to make cross-cultural comparisons on psychological constructs (F. Chen, 2007). However, in practice it is often an unrealistic goal, which is precisely why the alignment method was developed. A number of approaches are used to assess measurement equivalence, including multiple group confirmatory factor analysis (MGCFA: Bollen, 1989; Steenkamp & Baumgartner, 1998), item response theory analysis (IRT: Jilke et al., 2015), DIF-via logistic (Zumbo, 1999), and latent class analysis (LCA: Kankaraš et al., 2011). MGCFA is perhaps the most common method and has been used frequently in political science to assess cross-country equivalence of human values such as attitudes towards democracy and welfare (Ariely & Davidov, 2011), social and political trust (Freitag & Bauer, 2013), and national identity (Davidov, 2009).

Within the MGCFA approach, a number of forms of invariance have been proposed that one should test for, such as configural/form invariance (items load onto the same factors across heterogeneous groups); factor-level/metric invariance (where the loadings are roughly equal); and the intercept/scalar level (where the scores on the construct have the same unit of measurement and the same origin). Traditionally, scalar

invariance is required for comparison of latent means. If the factor loadings are equal across groups, metric equivalence is attained, and a one-unit increase would mean the same for all groups. A higher level of equivalence is assessed using "scalar variance", which constrains the factor loadings and intercepts to be equal across groups, and implies that respondents with the same value on the latent construct have the same expected response on the scale, regardless of their group (Davidov et al., 2015). Steenkamp and Baumgartner (1998) hold that configural and metrics invariance are insufficient to ensure valid comparisons of latent means across cultures, and accordingly, scalar invariance is necessary.

The challenges with cross-national comparisons on sociological scales is well reported. Perhaps differences in mean-values are artifacts of the scale construction/implementation and not "true" differences? Using the European Values Survey (EVS) and the European Social Survey (ESS), Adam (2008) studied cross-cultural social capital and notes that even within the same country in the same year, mean values of generalized trust differed by groups. They attribute this to differentiated scaling of the ESS and wording effects, and finds that the overall country ranking remains stable. Freitag and Bauer (2013) also found evidence that social trust was stable across cultural backgrounds using MGCFA. Based on an analysis of French, Italian, and German communities in Switzerland, they found that diversity of ethnic background does not constitute a barrier for analysis of social trust at the sub-national level.

Alignment Method

Traditional approaches to assessing cross-cultural measurement equivalence are overly strict, and "approximate" or partial equivalence might be sufficient (Davidov et al., 2015; Marsh et al., 2018; B. Muthén & Asparouhov, 2012; Van De Schoot et al., 2013). Partial equivalence implies that at least two parameters are equal across groups (e.g. loadings for partial metric equivalence and loadings plus intercepts for partial scalar equivalence), which allows for anchoring and DIF correction. Davidov et al. (2015) use the ESS to assess anti-immigration attitudes, and find approximate scalar equivalence in all ESS rounds, and that the "exact" approach is overly strict; which is when factor loadings and intercepts are assessed to see if they are identical across groups. The difficulty of achieving cross-cultural equivalence often results in researchers picking one of two problematic approaches: either a model with imposed equality constraints that fits the data poorly; or a model that fits the data well with no equality constraints. The resolution of this tension results in the third approach: which is to use approximate Bayesian measurement equivalence (Van De Schoot et al., 2013).

The use of the Bayesian approach involves the use of priors on specific parameters. In this approach, the average difference between loadings and intercepts across countries is assumed to be zero; like in MGCFA when testing exact measurement equivalence. However, the Bayesian approach permits small variations between parameters, whereas the exact approach constrains them to be equal. Van de Schoot et al. (2013) show that variance ≤ .05 imposed on the difference between intercept loadings does not result in biased conclusions. The Bayesian approach otherwise has similar constraints to the exact approach including: constraining the loading of one item to

exactly 1 in all groups and the intercept of the item is *exactly* 0 in all groups. Likewise, latent means and variances are freely estimated across groups. The Bayesian approach of approximate invariance "cannot establish approximate invariance when measurements are completely different . . . However, it can inform researchers when measurements are sufficiently similar to allow meaningful substantive comparisons" (Davidov et al., 2015, p. 262). Recent work by Marsh et al. (2018) seeks to extend the alignment approach within a SEM framework to develop a more flexible use of alignment with additional predictors and to compare model fit.

Logistic Regression DIF

Similar to the assessment of metric/scalar invariance with the alignment method, the detection of DIF via OLR signals that items are functioning differently across groups under investigation (Oliveri, Olson, Ercikan, & Zumbo, 2012; A. Wu, Li, & Zumbo, 2007; Zumbo, 1999). In the case of uniform DIF, one adjusts for overall test scores in the investigation of individual items, since particular group(s) might systematically benefit from the cumulative effects of DIF; resulting in overall higher test-level scores.

Methods

Measure

PISA is a math, science, and reading test administered every three years to 15-year old students. The most recent round of PISA surveyed 72 countries in 2015; data was released in December 2016. One of the explicit aims of PISA is to provide comparable data across OECD countries and other participating countries in order to improve educational policies and outcomes. In addition to the academic domains, PISA

collects information about material conditions for student learning and a variety of social-emotional and well-being questions that are relevant to educational policy and outcomes.

In total, six items (see Table 3.1) make up the PISA "Sense of Belonging to School" (Belong) scale. The responses to each item are on a 4-point scale: Strongly Agree (1), Agree (2), Disagree (3), Strongly Disagree (4).

Table 3.1 PISA Belonging Items

	Thinking about your school: to what extent do you agree with the following statements?
Item1	I feel like an outsider (or left out of things) at school.
Item2*	I make friends easily at school.
Item3*	I feel like I belong at school.
Item4	I feel awkward and out of place in my school.
Item5*	Other students seem to like me.
Item6	I feel lonely at school.

*Items were reverse coded to reflect a higher response scale (4) = an overall higher sense of belonging.

Sample Selection

The field of psychometrics has a well-documented literature about the difficulty of maintaining meaning across languages, which often results in DIF (Oliveri & Ercikan, 2011; Oliveri et al., 2012). Given this, I limit our analysis to students who took PISA in English, which is the largest language group in order to prevent any confounding effects of language and country or migration background. Roughly 83,300 students from 14 countries took PISA in English. Sample sizes by country can be found in Appendix Table 3.A. Likewise, in order to run the alignment method with a 6-item scale, there is a required minimum of 6 (unweighted) respondents per group (migrant background by country) in order to fit the model, with a suggested minimum of 20 respondents (L. Muthén, personal communication, May 24, 2018). As such, two countries (Hong Kong and Luxembourg) were excluded from analysis due to insufficient cell sizes and are not a part of the aforementioned 14 countries.

For both our analyses, I investigate cross-group comparability of students' immigration background: 1) Foreign-born (fb); 2) Native-born with foreign-born parents (nb2fb); and 3) Native-born with one foreign-born parent (nb1fb); and 4) Native-born with two native-born parents (heritage). Immigration background,

including nativity, duration of stay in a host country, and generational status is known to impact quality and even duration of life (Hummer et al., 1999; Portes & Rumbaut, 2001; E. J. Williams, 1995). All analysis is conducted with respect to country of PISA administration using PISA sampling weights. For the alignment, I have 56 groups (14 countries by 4 migration groups), which allows for both between- and within-country differences. It is worth explicitly stating that although the alignment is "more liberal" when compared to MGCFA by allowing for slight variance of parameters, this is still a strict assumption to test; that all 56 groups are treated as independent and are comparable. For OLR, I use country stratification with the four migration groups, which allows us to control for between-country effects and more closely assess the role of immigration background.

Analyses

Like a Bayesian Structural Equation Model (BSEM), which focuses on measurement parameters as approximate and the treatment of all parameters as variables, the alignment approach goes beyond the requirement of strict invariance in a CFA framework. The primary distinction for the alignment is the absence of *a priori* equality restraints on the factor loadings and intercepts across groups as suggested by:

$$\gamma_{\text{ijg}} = \nu_{jg} + \sum_{l=1}^{L} \lambda_{jlg}\, \eta_{ilg} + \varepsilon_{ijg} \quad (1)$$

γ_{ijg} defines the model for each ordinal item where i=1 ..., i, and i is the number of people, j =1 ..., j, and j denotes the items, g = 1, ..., g and g denotes the groups. $\nu_{jg}$is the intercept parameter, and $\lambda_{jlg}$is the loading parameter, η_{ilg} is the latent variable of person i, in group g, on factor l, and ε_{ijg} is the residual. The method begins by

estimating a model with a factor mean (α) set to 0, and factor variance (Ψ) set to 1 for all groups. From the initial model, all intercepts and loadings are estimated as free and unequal. For a technical description of the alignment method see Asparouhov and Muthén (2014) and for a descriptive examples of the alignment see Byrne and Van de Vijver (2017).

I ran the alignment method using maximum likelihood estimation with robust standard errors (MLR) and "free" group estimation, which is shown to work best with moderate to large degrees of invariance (Flake & McCoach, 2018). A total of 56 groups (4 migration groups by 16 countries) are used in the analysis. MPlus code can be seen in Appendix 3.C.

As an initial analysis to OLR estimation, I examined the internal structure of the PISA belonging scale. According to the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014), there are five sources of validity evidence for evaluating inferences from test scores: 1) test content; 2) international structure using factor analyses; 3) relationship to other variables (convergent/discriminant validity); 4) response processes; and 5) consequences of testing. Not all evidence is necessary or appropriate in all cases, however internal structure is fundamental to determine for purposes of dimensionality for test scoring. Likewise reliability estimates are important to understand the magnitude of effects for further analyses (Zimmerman & Zumbo, 2015). I examined whether the belonging scale fit a strictly unidimensional model by country, and, if not, whether essential unidimensionality is supported. Strict unidimensionality was determined based on a one-factor confirmatory factor analysis

(CFA) using MPlus. The four-point item responses were treated as ordinal using a weighted least square mean adjusted average (WLMSV) estimator. For a detailed description of the benefits of WLMSV estimators over Maximum Likelihood estimators with ordinal data see Beauducel and Herzberg (2006). Using Hu and Bentler's (1999) suggestions, a Comparative Fit Index of ≥ .95 and an RMSEA of <.08 were used to determine fit. I also considered a Tucker-Lewis Index of >.95. If strict unidimensionality was rejected, I assessed essential unidimensionality using exploratory factor analysis (EFA) by examining the ratio of first to second eigenvalues; if the ratio is > 4.0, then there is evidence for essential dimensionality (Slocum-Gori & Zumbo, 2011). The PISA belonging scale is expected to be unidimensional based on OECD analyses.

OLR uses the item response, regressed on adjusted-total score (belonging total score minus the item under investigation), and group membership. Items with $\chi 2$ values statistically significant at $p < .05$ for the grouping variable indicate the presence of DIF. Given that there are four groups, for each item in our DIF analysis, I ran a chi-squared test with three degrees of freedom (conditioned on the corrected total score) to investigate detection of DIF for at least one bi-variate comparison. If DIF is detected, I then examine all six bi-group comparisons; akin to the logic of a post-hoc test with one degree of freedom. Bi-group comparisons were made using confidence interval using STATA14's "pwcompare" command. Bi-group comparisons were only made on items with a significant omnibus test purely as an exploratory approach to understand the sources of DIF. Likewise, given the small number of items (6) and the importance of maintaining the integrity of the scale with the alignment method, even if there is some

evidence of DIF, I do not drop items for a subsequent analysis of the remaining items in the scale.

Both methods have advantages in the investigation of DIF. The alignment method allows for the simultaneous evaluation of DIF across many groups, which would otherwise be cumbersome in an OLR framework using contrast matrices. However, it is important to stress, the goal is not to assess the amount of invariance at the within- or between level, but to consider the role of immigration background in an international context. Likewise, the alignment method allows for *slight* amounts of measurement noninvariance in the comparison of many groups in a relative sense, whereas OLR is a more absolute comparison. On the other hand, OLR offers the advantage of investigating DIF on a construct while also controlling for overall levels of the construct using the item-corrected total. Moreover, the alignment method is used for group comparison using the entirety of the scale, whereas OLR allows a subtler investigation of where DIF may be emerging *within* the survey.

Results

Alignment

Full results of the estimated (non)invariance of the intercept and slope parameters can be seen in Appendix Table 3.C and Table 3.D respectively, and a summary of results is presented below in Table 3.2. Table 3.2 shows the number of noninvariant parameters by country and migration background. Model estimates were replicated. Except for Canada, all countries have at least one noninvariant parameter, which is perhaps expected for a construct that by definition depends upon one's social

context. However, consistent with the theory of alignment, the presence of a few noninvariant parameters does not mean that group comparisons are impossible. There is no established rule to say how much invariance should be tolerated; yet guidance by Asparouhov and Muthén (2014) is that no more than 25% of parameters should be noninvariant. Accordingly, for this analysis, if four or more parameters are noninvariant by migration group within country, then that group is considered not comparable to other groups. Totals that exceed the threshold of acceptability are bolded in Table 3.2.

Table 3.2 Number of Noninvariant Intercept and Slope Parameters by Country and Migration Background

	Intercepts				Factor Loadings			
Country	heritage	nb 1fb	nb 2fb	fb	heritage	nb 1fb	nb 2fb	fb
Australia	3					**6**		
Canada								
Ireland	**6**	2		1	**6**			
Lebanon	**11**	3		1				
Macao			2				**4**	1
Malta	**5**	3		3	**6**	**6**		
New Zealand	2							
Qatar			**7**	**7**		1	**6**	
Singapore	3	1		1				
Sweden		1						3
Trinidad and Tobago	**13**	3		**5**	**6**	3		
United Arab Emirates	**7**	1	**11**	**7**		1		
United Kingdom	**6**	1		1				
United States	1				**6**			
Total	57	15	20	26	24	17	10	4

Note: bold and underlined groups are considered noninvariant and are not comparable to other groups (percent noninvariant ≥ 25%)

For all countries except Canada, New Zealand, Singapore, and Sweden there is at least one incomparable group on either the intercept or slope or both, yet many groups are comparable. Invariance of the factor loadings can be understood "as the difference in factor score calibration with regard to the unit of measurement" (A. Wu et al., 2007, p. 10). This means that for those groups in countries with noninvariant loadings (Australia (nb1fb), Ireland (heritage), Macao (nb2fb), Malta (heritage & nb1fb), Qatar (nb2fb), Trinidad and Tobago (heritage), and USA (heritage)), a one-unit increase in "belonging" may not substantively or quantitatively mean the same thing. There is no clear pattern to these groups, except that most are heritage groups.

Substantively, the alignment results suggest that means of the construct are not comparable for those groups in countries with noninvariance of intercepts: Ireland (heritage), Lebanon (heritage), Malta (heritage), Qatar (nb2fb & fb), Trinidad and

Tobago (heritage & fb), United Arab Emirate (nb2fb & fb), and UK (heritage). It is perhaps unsurprising that the "heritage" group may often have a different conceptualization of belonging that includes personal family history. Likewise, for foreign-born persons and those born to foreign-born parents in Qatar and UAE, there are significant structural barriers to one's access to nationality that may influence their sense of belonging; even from a young age. And interestingly, in all three of the "settler" countries included in the sample (Canada, USA, New Zealand, and Australia) there is evidence of measurement invariance across all migration backgrounds, which is consistent with the narratives of being a "country of immigrants." Take the factor loadings of Ireland and Lebanon for a moment in the alignment analysis. For both countries it is the "heritage" group that is the only noninvariant group. If I assume for a moment that all foreign-born persons are Canadian, the results would suggest that Canadians (as immigrants) are comparable in both Ireland and Lebanon, but that it is the cross-country differences between those people from and within Ireland and Lebanon respectively that are noninvariant.

I also examined the number of noninvariant parameters by item (rather than migration group) as shown in Table 3.3. While no item meets the threshold of 25% to be considered noninvariant, two items (3 and 5) are approaching levels of concern.

Table 3.3 Noninvariant intercept parameters by item

Item1	17
Item2	14
Item3	28
Item4	13
Item5	28
Item6	18

Note: Each item has a total of 168 parameters estimated, thus an item would need roughly 33 non-invariant parameters to be considered non-comparable.

Flake and McCoach (2018) find that large amounts of noninvariance result in decreased true and estimated factor mean correlations and bias in factor means. This is an important consideration with these results, particularly with an emphasis on the "heritage" group; namely, that estimated means are likely biased for this group in comparison to those with an immigration background. This provides further justification for our OLR analysis to hold cross-country effects stable.

OLR Results

Results from our initial analysis are reported in Appendix Table 3.B. Estimated alphas are generally consistent with those reported by the OECD and range between .65-.87, or slightly higher in this sample likely given case selection of only English test-takers and the exclusion of individuals without an identified migration background. Factor analyses confirm strict unidimensionality of the overall scale; with essential unidimensionality for English speaking countries and Sweden and a 2-dimensional but highly correlated factors for Malta and all other countries. This suggests some cross-country differences of the scale dimensionality. Given that the OLR model tests for item

DIF by migration background with the pooled sample (using country stratification and weights), the factor structure is acceptable for further analyses.

OLR results suggests the presence of DIF on three of the six items (see Table 3.4: items 2, 3, and 5). These items are all those in which there is a "positive" frame on the statement (e.g. "I feel like I belong at school"), and no DIF was detected on "negatively" framed items (e.g. "I feel lonely at school"). Recognizing that the factor analysis confirms a one-factor solution; these results suggest that there is a potential wording effect on the PISA belonging scale.

Table 3.4 Ordinal Logistic Regression Results and Chi-square test by item

	item1	item2	item3	item4	item5	item6
	beta / (se)	beta / (se)	beta / (se)	beta / (se)	beta / (se)	beta / (se)
nb 1fb	0.02	-0.13	0.02	-0.07	-0.09	0.11
	(0.08)	(0.09)	(0.07)	(0.08)	(0.09)	(0.08)
nb 2fb	-0.12	**-0.21****	**0.24*****	-0.09	**-0.27*****	**0.14****
	(0.06)	(0.07)	(0.07)	(0.07)	(0.08)	(0.07)
fb	0.06	-0.15	**0.35*****	0.02	**-0.34*****	0.00
	(0.07)	(0.08)	(0.07)	(0.08)	(0.10)	(0.08)
AdjTot	0.67***	0.86***	0.75***	0.92***	0.80***	1.01***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
cut1	6.45***	8.08***	7.46***	9.27***	6.38***	10.25***
	(0.16)	(0.20)	(0.18)	(0.22)	(0.18)	(0.23)
cut2	8.89***	11.42***	10.34***	12.37***	9.24***	13.04***
	(0.18)	(0.22)	(0.21)	(0.26)	(0.22)	(0.27)
cut3	12.21***	15.97***	14.50***	16.52***	15.32***	17.14***
	(0.21)	(0.27)	(0.25)	(0.30)	(0.29)	(0.32)
Test- chi2(3)	4.78	**11.76**	**31.62**	2.27	**20.07**	5.56
Prob > chi2	0.19	0.01	0.00	0.52	0.00	0.14
Effect size of migback	0.001	0.165	0.161	0.135	0.155	0.139

Note: **Bold and underlined** parameters indicate detection of group DIF. "Heritage" is the referent group. Groups are: native born with 1 foreign-born guardian (nb 1fb); native born with 2 foreign-born guardians (nb 2fb); and foreign-born (fb). "AdjTot" is the Adjusted-Total Score. Cut scores represent the threshold cut for the 4-point Likert scale. Effect size of migback is the difference in R_2 values with the inclusion of the migration background contrasts.

For all items, I conducted a chi-square test with three degrees of freedom to compare the four migration groups. For items 2, 3, and 5, the test is significant at the 99% confidence level, and I include the effect size of the change in R_2 with the inclusion of migration background in the model per item. To investigate the sources of DIF for these items, I performed a pairwise comparison of all groups for each item (see Table 3.5. below). The contrast is read as how much higher (or lower) the second group is compared to the first, and those comparisons whose confidence interval does not contain 0 are considered meaningfully different and are bolded below. A total of five out eight of the significant bi-group comparisons are with the "heritage" group, which is consistent with alignment results.

Table 3.5 Pairwise comparisons of DIF items

	Contrast	SE	95% Conf. Interval	
item2				
nb with 1 fb parent vs heritage	-0.13	0.09	-0.30	0.03
nb with 2 fb parents vs heritage	**-0.21**	**0.07**	**-0.35**	**-0.07**
fb vs heritage	-0.15	0.08	-0.31	0.02
nb with 2 fb parents vs nb with 1 fb parent	-0.08	0.10	-0.28	0.12
fb vs nb with 1 fb parent	-0.02	0.11	-0.23	0.20
fb vs nb with 2 fb parents	0.06	0.10	-0.14	0.26
item3				
nb with 1 fb parent vs heritage	0.02	0.07	-0.12	0.16
nb with 2 fb parents vs heritage	**0.24**	**0.07**	**0.11**	**0.37**
fb vs heritage	**0.35**	**0.07**	**0.21**	**0.50**
nb with 2 fb parents vs nb with 1 fb parent	**0.22**	**0.09**	**0.05**	**0.40**
fb vs nb with 1 fb parent	**0.33**	**0.10**	**0.15**	**0.52**
fb vs nb with 2 fb parents	0.11	0.09	-0.07	0.29
item5				
nb with 1 fb parent vs heritage	-0.09	0.08	-0.25	0.08
nb with 2 fb parents vs heritage	**-0.27**	**0.08**	**-0.42**	**-0.11**
fb vs heritage	**-0.34**	**0.10**	**-0.54**	**-0.14**
nb with 2 fb parents vs nb with 1 fb parent	-0.18	0.11	-0.39	0.03
fb vs nb with 1 fb parent	**-0.26**	**0.12**	**-0.50**	**-0.01**
fb vs nb with 2 fb parents	-0.08	0.12	-0.31	0.16

Overall the alignment method and OLR suggest it is *inappropriate* to treat all peoples of different immigration backgrounds as equivalent on the PISA measure of belonging. Migration backgrounds within the 14 countries are not exchangeable, and one's background is a relevant factor to consider when thinking about an individual's interaction with the items of the PISA belonging scale.

For both methods, results suggest that in most countries it is the "heritage" group parameters that are not comparable to others with some kind of migration background. Said another way, there seems to be more "within-group" variability for native-born persons with native-born parents in their respective PISA test-taking country than compared to non-heritage students. In one way this is unsurprising since this group is significantly larger. However, given the wide variety of cultural backgrounds students with an immigration background are coming from, the "within-group" variability of the "heritage" populations is striking. Indeed recent literature stresses increased attention be paid to the unique interaction of national culture of the host society with the disparate cultures of incoming migrants (Alba & Foner, 2014; Koopmans et al., 2005; Wright & Bloemraad, 2012).

Discussion

The two techniques in this paper confirm our hypothesis about the presence of DIF based on migration background. Analyses highlight the way in which "heritage" students have the most noninvariance, which is suggestive of between-country variance. This is generally unsurprising when considering immigration, given that one's

integration and sense of belonging is connected to their receiving country. Interestingly, the difference cannot simply be attributed to one's nativity and rearing within a particular country, since three of the groups are "native-born" students. Yet, with great caution, results from the alignment analysis suggests that indeed for many groups across countries it is possible and permissible to compare measures of social belonging using the PISA. However, further investigation should be warranted to consider the number of noninvariant parameters and groups considered acceptable when using and interpreting results from the alignment method.

Both analyses confirm that migration background is an important variable to consider as a part of the ecology of student testing and measures of quality of life. Yet, consistent with past research (Y. Chen, 2017), the OLR results suggest that one should be cautious in their use of negatively-keyed and negatively worded items, and that the presence of negatively framed items may have an effect on DIF results. Assuming the OLR results are related to the construct framing at the item-level, these results could feasibly have an impact on the results of the alignment method; which examines the groups on all items simultaneously. Interestingly, results from the alignment method seem to capture the differential functioning of items 3 and item 5; however, the procedure is still not sensitive enough to flag these items as DIF compared to logistic regression, and does not capture DIF at all for item 2. An important caveat to cross-country comparisons concerns language adaptation. Future research is needed to determine the extent to which the alignment method can tolerate minor degrees of variance due to testing language.

The alignment method as put forward by Muthén and Asparouhov (2012) was introduced in the research literature with binary response data, and Flake and McCoach (2018) conducted a simulation study to offer an exploratory justification for the use of alignment with Likert-style response options. As such, this research offers an important extension of the alignment method into social-science research using real-life application of the alignment with ordinal, Likert-style response data and using complex survey data. Likewise, that these authors are aware of, this paper present the first multi-group, DIF analysis of Social Belonging upon immigration background.

For immigrants, the idea of "belonging" is rarely studied within a methodological sense and has been largely ignored in the assessment field. Instead, immigrant belonging is tied up within the broader conversation of "integration." Perhaps obviously, this research is tied up with the shift towards "civic" nations away from "ethnic" nations; where anyone can integrate into the common culture. Yet, cultural consideration is important given its role to access and understand a multitude of meaningful ways of living and knowing (Dworkin, 1985). We need to give serious methodological consideration to cultures in the way we construct, analyze, and use measures; particularly when these measures are tied to one's deeper ecological setting. Human learning and development are situative and multiplicative (Nasir, 2008; Nasir & Saxe, 2003), and it behooves us to develop richer theory that can accommodate people's ecological settings, including migration background. By taking into account the measurement invariance of migration background, policy conversations are improved with more nuanced consideration of peoples' contexts. Likewise, measurement theorists

should also consider how migration background should be taken into account not just in terms of measurement invariance, but the way in which one's migration background actually (re)shapes the constructs at hand. Individuals do not live in isolated environments, and with increased migration we all collectively need to rethink what it means to belong and how we can foster social cohesion.

Chapter 4: Multicultural Integration Policy as an Explanatory Factor of Immigrant Social Belonging: Multilevel Evidence for a Multilevel Construct

Introduction

Large scale international assessments, like the Program for International Student Assessment (PISA) are useful tools to help track social and educational outcomes of immigrants in comparison to non-immigrants within a country. PISA is a math, science, and reading test administered every three years to 15-year old students in participating countries with the explicit use of rating/ranking country-level education systems to inform policy development. This ranking implicitly invokes a multilevel framework as inferences are made about countries from students nested within schools. This paper seeks to do two things. First, I provide multilevel evidence to contribute to the validation of PISA's social belonging measure with consideration for one's immigration background. Second, I argue that country-level integration policy can be considered as an explanatory variable to inform the validity argument of social belonging using an ecological framework; which has largely been ignored in psychometric literature.

PISA aims to provide comparable data across the Organization for Economic Cooperation and Development (OECD) and other participating countries. In addition to the academic domains assessed, PISA asks a variety of social-emotional/well-being questions that are relevant to educational policy and outcomes. As defined by Forer and Zumbo (2011), and Zumbo and Forer (2011), the PISA measure of social belonging is a multilevel construct; meaning it is a phenomenon that is potentially meaningful at the individual level and at least one level of aggregation; however the primary interpretation and use is at the aggregate level. This is evidenced by the de facto use of PISA to compare countries and support school-level

policy using aggregated, student-level data. Student-level reliability data cannot be assumed to signify reliability of measurement claims at the aggregate level (Zumbo & Forer, 2011). However, there is little research about the appropriateness of making aggregate-level claims on measures like PISA, and no research to date that this author is aware of concerning the belonging scale used in PISA. Likewise, no research has yet been done to connect policy indicators to student-level outcomes with respect to immigration background.

PISA has the explicit goal to evaluate past and existing country policies and to inform future policy development consistent with the OECD motto: "Better Policies for Better Lives." Not only are national institutions assumed to be important to craft education policies, but national characteristics are pervasive across all aspects of denizen's lives consistent with the ecological model of item response as put forward by Bronfenbrenner (1981), Zumbo et al. (2017), and others. To paraphrase the renowned population health scientist Clyde Hertzman, national policies shape the social world in a way that "gets under their skin" and affects trajectories of human development (Hertzman & Boyce, 2010).

I present multilevel validation evidence of PISA's social belonging measure. Then, I conduct a series of multilevel models that connect PISA with policy indicators, which are explicitly designed to target immigrants and families of immigrants to assistant in the integration process. Roberson and Zumbo (2019) find that immigration background is an important consideration when examining social belonging. As such, this research extends the scholarly understanding using an empirical example of how policies designed to ease the integration process for immigrants relate to one's sense of social belonging.

Background

Belonging

Social belonging, defined as "a sense of having a positive relationship with others" (Baumeister & Leary, 1995) is a central human need, and the lack of belonging, or exclusion, harms people in a number of ways including, subjective well-being (Lyubomirsky et al., 2005), intellectual achievement (Walton & Cohen, 2007), health and immune functioning (Berkman & Syme, 1979), and self-control (Baumeister et al., 2002). Belonging entails having the opportunity to participate in the design of social and cultural structures, and having the right to contribute to, and make demands on, society and political institutions (powell & Menendian, 2006). For powell and Menedian (2006), belonging stands in contrast to "othering" which they define as the "set of dynamics, processes, and structures that engender marginality and persistent inequality across any of the full range of human differences based on group identities" (pg. 17). Belonging is both a personal quality and reflects how groups are positioned/received/regarded in society and reflects both objective qualities of power and resources and the intersubjective nature of group-based identity. When aggregated, individual-level acts of discrimination and marginalization have a magnifying effect on group-based inequalities (Blank, 2005). As such, belonging is inherently political and necessarily concerned with both the personal and the aggregate in a multilevel way. Indeed, only by considering belonging within a multitiered framework can we address group-based marginalization and inequality. The most important good society distributes to persons is membership (Walzer, 1984). "The right to belong is prior to all other distributive decisions since it is members who make those decisions" (powell & Menendian, 2006, p. 35). Meanwhile access to society is not always sufficient to create

integration: some groups require special accommodation. Through a lens of pluralism and multiculturalism, society can not only accommodate differences, but support new inclusive narratives, identities, and structures (Redhead, 2002).

Multilevel validation

"Multilevel validation methods aim to provide a strong form of construct validity; that is, the evidence should provide an explanation for the observed variation in test scores" (Zumbo et al., 2017, p. 342). Multilevel validation is more than analyzing multilevel data, but offers an explanatory model about the construct under investigation. I analyze the PISA belonging scale to understand the cross-national variation in social belonging and its connection to immigration background using an explanatory model of integration policy.

Like other assessments examined within a multilevel framework, PISA is explicitly designed to provide feedback at an aggregate (school/region/country) level and is not intended to provide any feedback for the individual students. It is used to inform policy of educational systems and larger social policy. Given this, social belonging should be interpreted within a multilevel ecological model. Within the context of educational settings most constructs are inherently multilevel and methodological consideration should be assumed (Klein et al., 1994). Given the use of multilevel constructs to shape policy in schools and countries for all the target population in a "high stakes" context, it is important that adequate conceptual and methodological thought is given to the evaluation of these measures.

Questions about belonging are asked of individuals, who consider their "group-belonging" when understanding one's individual sense of belonging (G. L. Cohen & Garcia, 2005). In short, this requires an assessment of the level of theory in order to explain data variability in multilevel validation (Zumbo & Forer, 2011). Since the level of measurement of belonging is at the student-level within schools, and the inferences are being made about groups within countries, there is enormous potential for spurious claims, and as such our analysis needs to properly consider the structure of our data. More specifically, we ought to avoid the "atomistic fallacy", where unjust inferences are drawn at the aggregate level from individual level data; and the ecological fallacy, where unjust inferences are made about individuals from aggregate level data (Bliese, 2000).

Within the context of multilevel research, Chen, Matthieu, Bliese, Yammarino, and Dansereau (2004) propose six compositional models. The selected-score model uses a single, individual score to define the group-level construct. Second, the summary index model uses a statistic (such as mean or sum score) of individual scores to define the group. Third, the consensus model which measures within group agreement using items about a lower level (e.g. individuals rate their own group motivation). Fourth, the referent-shift consensus model which measures within group agreement about items pertaining to the group. Fifth, the dispersion model measures within-group heterogeneity (e.g. differing skills in a team). Then the sixth, aggregate properties model measures constructs directly at the group-level (e.g. supervisor rates an entire team's productivity).

Once the framework a multilevel construct is understood, one can investigate the patterns of variability across levels to draw conclusions. A developed approach is the within- and between- group analysis of variance, called WABA (Da Costa & Araújo, 2012; Dansereau & Yammarino, 2000). Groups can be composed of homogeneous, independent, or interdependent members; referred to as wholes, equivocal, and parts, respectively. If group members are homogenous, a single value can describe the entirety of the group; which assumes that only between-group variation exists in the data/construct under investigation. Conversely, with independent members, all variation is assumed to be at the individual level. Lastly, the parts view suggests there is individual variation, but it is dependent upon one's group context.

To make inferences using WABA, Dansereau and Yammarino (2000) propose two statistics to be used: E-ratio to assess practical significance and the F-ratio to assess statistical significance. One should first confirm the F-ratio before assessing the inference of the E-ratio. The F-ratio is the typical Mean Square(between)/Mean Square(within) when a "wholes" view is implied, and is the inverse ratio for "parts" view. The E-ratio is the between-group eta correlation divided by the within-group eta correlation. When within- and between- group variation is equal, the E-ratio is 1[3]. The E-ratio has a lower bound of 0 (no variation between groups), and no upper bound. For both ratios, one or more predetermined critical values need to be set before making inferences (Dansereau et al., 1984). The 15° test divides the overall variance into three

[3] This is equivalent to Cohen's f effect size statistic.

equal parts, while the 30° test creates equal intervals in terms of angles and is more conservative. The 30° test requires within- or between- group variation to exceed 75% of the total before an inference about wholes or parts may be made.

Methods

Data

PISA is administered every three years to 15-year old students in participating countries. The most recent round of PISA administration surveyed 72 countries in 2015, and the data was released in December 2016. Each item is on a 4-point scale (Strongly Agree, Agree, Disagree, Strongly Disagree), recoded 0-3, with 3 reflecting a higher level of sense of belonging, and three items (1, 4, 6) are reverse coded so that a higher score represents a higher sense of belonging. For purposes of the total score, the sum is taken; with a possible range between 0-18. Sample sizes by country (and language) can be found in Appendix Table 4.A.

Table 4.1 PISA Belonging in School Items

Item	Thinking about your school: to what extent do you agree with the following statements?
ST034Q01TA	I feel like an outsider (or left out of things) at school.
ST034Q02TA*	I make friends easily at school.
ST034Q03TA*	I feel like I belong at school.
ST034Q04TA	I feel awkward and out of place in my school.
ST034Q05TA*	Other students seem to like me.
ST034Q06TA	I feel lonely at school.

Note: * items reverse coded to reflect a higher score is a higher sense of belonging

In addition to PISA, I include country-level policy indicators using the Migration Integration Policy Index (MIPEX). MIPEX covers eight policy domains: labour market mobility, family reunion, education, political participation, long-term residence, access to nationality, anti-discrimination, and health. These eight domains are aggregated based on 167 policy indicators (see Appendix Table 4.B. for MIPEX values and domains by country). For each policy indicator, a maximum of three points can be awarded; where a score of three means the highest standard for equal treatment is met between immigrants and non-immigrants. For the sake of rankings and comparisons, the 1, 2, 3 scale is converted to a 0, 50, 100 scale; where a 100 is the highest value per domain. Finally, within each domain, the policy indicators are considered separately within four dimension scores, which are subsequently averaged to get a single domain-score per country. It is important to stress how national policy is more than a mere descriptor of the state of affairs, but can also be perceived as a kind of intervention: where governments seek to shape trajectories.

The indicators are completed by national experts and are peer reviewed.[4] The inclusion of MIPEX over other policy indicators such as Citizenship Policy Indicator or the Multiculturalism Policy Index is for theoretical and practical reasons. Helbling (2013) demonstrates how many of the sub-domains of these policy indicators are highly-correlated, however as a whole, one cannot flippantly select an index given that the overall scores often reflect distinct subject areas. As such, policy index selection should be guided by substantive research questions. I am interested in integration policies (managing those already living within a country) more than immigration policy (determining who is permitted to enter the country). Likewise, as a part of the investigation into social belonging, I have an interest in exploring the various mechanisms that contribute to increased well-being above and beyond one's legal status (e.g. attaining citizenship or not). As such, I use the MIPEX domain areas and not the total score, which provides the most coverage of policy areas. More specifically, I use six of the eight MIPEX domains including: Family Reunification, Education, Permanent Residency, Access to Nationality, and Anti-Discrimination Policies. I exclude Labour Policy and Political Participation given the data linkage to 15-year-old students, who by merit of their age, are ineligible to vote and generally ineligible or unlikely to work. Likewise, the most recent MIPEX survey provides the best alignment of dates to match the PISA administration (2014 and 2015 respectively) and provides the largest coverage of countries (38 in all) compared to other indicators. It should be noted that MIPEX only

[4] For more information on MIPEX methodology, please visit mipex.eu

conveys what policies are *intended* to do, and these indices are additive in nature (Goodman, 2010); meaning within each domain, policy indicators are given equal weight. Between the PISA and MIPEX, there is an overlap of 36 countries, which is our final analytical sample representing over 265,000 students in 10,377 schools.

Analysis

As a preliminary analysis, I investigated to what extent one can make comparisons on the Belonging scale across different language groups. Differential item Functioning (DIF) based upon language is a well-documented phenomenon; see Oliveri and Ercikan (2011), Oliveri, Olson, Ericikan, and Zumbo (2012). I used MPlus and the alignment method using maximum likelihood estimation with robust standard errors (MLR) and "free" group estimation, which is shown to work best with moderate to large degrees of invariance (Asparouhov & Muthén, 2014; B. Muthén & Asparouhov, 2012; Flake & McCoach, 2018). The alignment approach of approximate invariance "cannot establish approximate invariance when measurements are completely different . . . However, it can inform researchers when measurements are sufficiently similar to allow meaningful substantive comparisons" (Davidov et al., 2015, p. 262). For this analysis, there are 31 language groups (see Appendix Table 4.A.)

After our preliminary language analysis, I conducted a series of Exploratory Factor Analyses (EFA) to investigate the use of the Belonging scale as a total score. I ran a single-level EFA without respect to country or language (1 or 2 factors); a multilevel EFA nested within country with 1 and 2 factors at the within and between-level; and an

identical multilevel model, treating the data as categorical rather than continuous. EFA analyses were completed using MPlus 8.

Central to measurement, and especially multilevel measurement, is the issue of reliability. Namely, individual reliability estimates cannot be assumed to be the same at the aggregate level (Bliese, 2000). Chen et al. (2004) suggest that for summary index models, one should consider the level of non-independence in the data using ICC(1), which is the proportion of individual variance influenced by the group belonging. Then, the within-group reliability, measured by the ICC(2) indexes the reliability of differences between group means by measuring the proportional consistency of variance across groups (Bliese, 2000). Collectively these two ICC measures are often referred to as a Within And Between Analysis (WABA) of variance. High within-group reliability is beneficial to allow for groups means to be estimated from a smaller sample from within the group. The WABA analysis was conducted in R using the WABA package (O'Connor, 2019).

Finally, after determining the appropriateness of comparison across language group and the sum score, I conduct a series of multilevel models in order to investigate the relationship between integration policy and students' sense of belonging in school. I use an iterative approach to model building for a total of 5 models; where each model is nested within the subsequent one. There are a number of trade-offs to consider when building and selecting a multilevel model including which parameters to free or make random, what interactions are used, the number of covariates, and the estimation method. For this analysis, the primary investigation is about country to country

differences ("level-3") and the way in which national-level, integration policy is related to individuals' sense of belonging at school with unique migration backgrounds (i.e. the interaction terms in Model 5). As such, I am not, at present, interested in differences between schools (either within or between countries), but focus on the cross-level interactions of policy by migration background. Nor am I directly interested in the comparison of peoples with the "same" migration background across countries; indeed, this in conceptually difficult to envision. Likewise, for computational reasons, it would be taxing and result in biased estimates to include countries and schools as fixed-effects in a model. So, while complex, the multilevel models provide the necessary flexibility to correctly account for cross-level variances in a reasonably efficient manner.

(1) The first model is the "null" model, which includes a dependent variable (Total Social Belonging) nested within school, nested within country, with no predictors or controls. (2) Model two, then introduces fixed language-family controls, which I refer to as the "baseline" model, since these parameters are not substantively interpretable, but are controls to account for variance attributable to language separate from either school or country. (3) Model three then introduces two "level-1" predictors: Female (coded 1 if female and 0 if Male) and three categorical variables accounting for migration background including "nb1fb" (native born with 1 foreign-born guardian), "nb2fb" (native-born with 2 foreign born guardians), "fb" (foreign born), and "heritage" (native born with 2 native born guardians) is our omitted/referent group. (4) Model four is theoretically optional. Here I introduce main-effects of the six policy domains at the country-level. Given our interest in the cross-level interactions with migration

background, this step is done to check the improvement of model fit with country-level variables. Finally, (5) in model five I investigate cross-level interactions of policy with the three student migration background variables above (fb, nb2fb, nb1fb, with heritage omitted). Immigration background, including nativity, duration of stay in a host country, and generational status is known to impact quality and even duration of life (Hummer et al., 1999; Portes & Rumbaut, 2001; E. J. Williams, 1995). MIPEX policies are explicitly designed to be tailored to newcomers within a country. As such, I do not expect the policies to have a significant effect for heritage students, and I would expect the greatest impact for foreign-born students, children of two foreign-born parents, and children of one foreign and one native parent in that order. Multilevel models were completed with STATA 14 using Maximum Likelihood (ML) estimation. Restricted-maximum likelihood (REML) estimation provides less biased estimates compared to ML with limited number of higher-level variables and better handles high-correlation of covariates. ML is more appropriate when data are unbalanced. Likewise, ML allows one to compare models with different fixed-effects comparisons. However, with very large samples the differences between REML and ML are minimal. For primarily theoretical reasons, I do not include population weights in our analysis. While necessary to draw inferences about individuals, the primary goal of our multilevel model is to evaluate the impact of policy frameworks on groups of students. As such, the inclusion of population weights would overly bias the estimated parameters in favour of larger countries like the United States and Germany. So, while there is already some bias

towards those countries with a larger sample, omitting country-weights more appropriately allows us to compare policy impacts across countries.

Results

Preliminary Results

Results (see Appendix Table 4.C.) from the alignment analysis on language confirm our hypothesis that there is measurement invariance based upon the language of test administration. 11 out of the 31 languages demonstrated sufficient levels of invariance to be considered comparable, which were principally Balto-Slavic and Finnic languages. However, given the high-levels of non-invariance across language groups, our results confirm the need to account for language of test-administration in subsequent analyses. Yet, there is not a simple pattern of language and country. Most countries only have a single language of test administration, while a number of countries administered the PISA in three or four languages (e.g. Spain, Belgium, Switzerland, Lithuania, Italy). Some languages only appear within a single country (Basque, Korean, Icelandic, Japanese, Turkish, Greek). Then some languages appear in many countries; for example, English is administered in 8 countries.

This creates a number of challenges for subsequent analyses. Namely, it is not possible to control for individual language and country-level effects due to language-country confounds (e.g. treated separately, the effects for Greek and Greece would be equal), likewise it would ill advised to include each language effect for a total of 30 additional language variables. Thus, while sub-optimal, I introduce language-family effects into our model to account for language non-invariance while investigating

country-level effects. I include a total of 5 language families: Germanic, Romantic, Balto-Slavic, Finnic, and an artificial "Singular." The "Singular" family consists of 5 languages (Greek, Turkish, Korean, Japanese, and Basque) which do not belong to a common family of any other language considered.[5] The language-group effects are not meant to be interpreted, but are controls to better estimate country-level effects, while reducing the confounding effect of language of instruction.

Multilevel-Validation Evidence

Our single-level EFA results (see Table 4.2. below) confirm "essential" unidimensionality of the Belonging scale. Hu and Bentler (1999) suggest a Comparative Fit Index of ≥ .95 and an RMSEA of <.08 be used to assess "strict" dimensionality; which was not attained. Since strict unidimensionality is rejected, I examined the ratio of the first to second eigenvalues consistent with Slocum-Gori and Zumbo (2011), where a ratio > 4.0 provides evidence of "essential unidimensionality." Of note, in both the single-level and multilevel (continuous) context, the 2-factor model appears to have a better fit. Upon examination of the item-loadings, the second factor appears to be those that are negatively framed. This suggests that there is a plausible "wording-effect" consistent with the results by Roberson and Zumbo (2019). Also, of important note, the multilevel EFA treating the items as categorical could not be estimated, likely due to limited variability at the country-level and the increased computational demands of an

[5] The relationship between Korean and Japanese is still contested. However, some scholarships suggest both may belong to a larger "Altaic" group which would subsequently also include Turkish.

ordinal model. Indeed, the ICC values generated in the multilevel (continuous) model suggest limited item-level variability at the country-level.

Table 4.2 Summary EFA Results

	Single-level EFA			Multilevel EFA continuous												ICCs
	1-factor	2-factor		1 within/1 between		2 within / 1 between			1 within/ 2 between			2 within/ 2 between				
Loadings		1st	2nd	W	B	W1	W2	B	W	B1	B2	W1	W2	B1	B2	
item1	0.81*	0.83*	0.00*	0.74*	0.90*	0.76*	0.00*	0.90*	0.72*	0.96*	0.00	0.76*	0.00*	0.96*	0.00	0.04
item4	0.80*	0.85*	-0.03*	0.73*	1.02*	0.77*	-0.01*	1.02*	0.73*	0.77*	0.27*	0.77*	-0.01*	0.77*	0.27*	0.03
item6	0.88*	0.88*	0.04*	0.80*	1.02*	0.83*	0.00*	1.02*	0.80*	0.7*	0.34*	0.83*	0.00*	0.71*	0.34*	0.04
item2	0.73*	0.00	0.821*	0.65*	0.61*	-0.00*	0.77*	0.61*	0.65*	0.00	0.99*	-0.00*	0.77*	0.00	1.00*	0.02
item3	0.66*	0.19*	0.52*	0.58*	0.31*	0.11*	0.54*	0.31*	0.58*	0.82*	-0.45*	0.11*	0.54*	0.82*	-0.45*	0.04
item5	0.71*	0.00	0.78*	0.61*	0.78*	-0.00*	0.71*	0.78*	0.61*	0.40*	0.49*	-0.00*	0.71*	0.40*	0.49*	0.03
Factor Correlations		0.76	-	-	-	0.73	-	-	-	0.54	-	0.73	-	0.11	-	
Eigenvalues				[within]	[between]											
First	3.83			3.36	4.10											
Second	0.77			0.85	1.11											
Third	0.52			0.59	0.43											
Fit Statistics																
RMSEA	0.18	0.09		0.78		0.24			1.75			0.87				
90% C.I	.18-.18	.09 - .09		.78 -.78		.24 -	.24		1.75 - 1.75			.87 -	.87			
CFI	0.96	1.00		0.73		0.98			0.01			0.85				
TLI	0.93	0.98		0.55		0.96			-1.28			0.44				
SRMR	0.07	0.02		0.07	0.13	0.02		0.13	0.07	0.04		0.02		0.04		

Note: (N=267821; K=36 [multilevel only]); WLSMV estimator; 4th-6th eigenvalues not reported; the Multilevel categorical EFA could not be estimated

I used our WABA (see Table 4.3.) analysis to further investigate the country-level variation using the Belonging total-score. An E-ratio of less than .77 (15-degree test) or .58 (30-degree test; which is more conservative) suggests a parts interpretation in terms of practical significance. Results suggest that the belonging scale has a "parts" inference with an E-ratio of .24; meaning the understanding of the construct is understood both at an individual level and country level. For statistical significance, the F-ratio is used, but contrary to the ANOVA case, the F-ratio in WABA is reversed. An F-value of less than one suggests the within-class effect is rejected. In our case I accept that there is a within-class effect.

Table 4.3 WABA Results

	Mean	SD	Eta-Correlation, between	Eta-Correlation, within	E-ratio	E-Ratio Inference	F-Ratio (within/ between)	Probability for F-Ratio	ICC(1)	ICC(2)
Belong Total	12.451	3.532	0.237	0.971	0.244	Parts	0.002	1	0.058	0.998

Multilevel Model Results

In total, our multilevel model included 265,189 students, in 10,377 schools, in 36 countries. Looking at the null model (#1) and models with only student-level variables (#2 & #3) in Table 4.4, between 4% - 5% of the total variance is accounted for at the country-level and roughly 6% - 7% is accounted for at the school-level. This is consistent with our hypothesis that the majority of variance of sense of belonging at school would be explained at the individual-level, then the school, then at the country-level. Yet, while unreported here, a likelihood ratio test of a two-level model (students in school) compared to the simple three-level model (#1: students in school in country) suggests that the inclusions of country as a nesting variable is a better fitting model. Related,

assessment of the model residuals (see Appendix Figure 1-4.1. for Model 5 Residuals using Q-Q plot) suggests that errors are normal and that multilevel models are appropriate.

Table 4.4 Multilevel Model Results

	Model 1	Model 2	Model 3	Model 4	Model 5
Fixed Effects					
Germanic		0.02	-0.00	-0.00	0.00
		(0.22)	(0.22)	(0.21)	(0.21)
BaltoSlavic		-0.64***	-0.65***	-0.65***	-0.66***
		(0.05)	(0.05)	(0.05)	(0.05)
Romantic		0.02	0.00	-0.01	-0.02
		(0.05)	(0.05)	(0.05)	(0.05)
Finnic		-0.48***	-0.52***	-0.52***	-0.54***
		(0.08)	(0.08)	(0.08)	(0.08)
Female			-0.29***	-0.29***	-0.29***
			(0.01)	(0.01)	(0.01)
fb			-0.54***	-0.54***	0.54**
			(0.03)	(0.03)	(0.26)
nb1fb			-0.14***	-0.14***	0.06
			(0.02)	(0.02)	(0.19)
nb2fb			-0.10**	-0.10**	-0.30
			(0.03)	(0.03)	(0.27)
Policies					
Family Reunification				0.00	0.00
				(0.01)	(0.01)
Education				0.00	0.00
				(0.01)	(0.01)
Permanent Residency				0.03**	0.03**
				(0.01)	(0.01)
Access to Nationality				0.01	0.01
				(0.01)	(0.01)
Anti-Discrimination				-0.01**	-0.01**
				(0.01)	(0.01)
Health				0.02	0.02
				(0.01)	(0.01)
Policy/Migration Interactions					
fbXfamily					-0.01**
					(0.00)
fbXEducation					0.01***
					(0.00)
fbXRes					-0.01**
					(0.00)
fbXNat					-0.01***
					(0.00)
fbXAnti-Discrim					0.01***
					(0.00)
fbXHealth					-0.01***
					(0.00)
nb1fbXfamily					-0.00
					(0.00)
nb1fbXEducation					0.00
					(0.00)
nb1fbXRes					-0.00
					(0.00)

	nb1fbXNat					0.00
						(0.00)
	nb1fbXAnti-Discrim					0.00
						(0.00)
	nb1fbXHealth					-0.00
						(0.00)
	nb2fbXfamily					-0.01**
						(0.00)
	nb2fbXEducation					0.01**
						(0.00)
	nb2fbXRes					-0.00
						(0.00)
	nb2fbXNat					-0.00
						(0.00)
	nb2fbXAnti-Discrim					0.01**
						(0.00)
	nb2fbXHealth					0.00
						(0.00)
	constant	12.28***	12.39***	12.60***	10.45***	10.40***
		(0.12)	(0.13)	(0.13)	(0.73)	(0.74)
Random Effects						
	School	-0.32**	-0.30**	-0.28**	-0.52***	-0.51***
		(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
	Country	-0.58***	-0.62***	-0.61***	-0.61***	-0.62***
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
	constant	1.22***	1.22***	1.22***	1.22***	1.22***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ICC	Country	.043	.045	.047	.029	.030
	School |Country	.068	.069	.071	.054	.054
LR Test Chi2		-	219.36	856.31	17.46	215.32
p-value of Chi2 test		-	0.000	0.000	0.008	0.000

Results of the main effects have several interesting findings. On average, females tend to report lower-levels of sense of belonging when compared to males, which is constant across all models. Consistent with our hypothesis in models 3 & 4, people with an immigration background (fb, nb2fb, & nb1fb) have statistically significantly lower levels of sense of belonging at school when compared to "heritage" students. Indeed, consistent with our expectations foreign-born students tend to have the greatest difference with heritage students, while the difference of nb2fb and nb1fb is

comparable. The introduction of policy*migration background interaction terms adds complexity to the interpretation, which is discussed below. While there is no well-defined effect size to evaluate cross-level interactions (Hox, Moerbeek, & Schoot, 2017), results are still worth exploring. Main effects of the multicultural policies (models 4 & 5), suggests that for all students, access to permanent residency has a positive and statistically significant relationship with sense of belonging. Meanwhile, contrary to our hypothesis there is a small, but statistically significant negative relationship between Anti-Discrimination policies and sense of belonging at school. Perhaps explained by the fact that if a country needs anti-discrimination policies, all denizens of the country are experiencing a critical lack of social cohesion.

Consideration of the policies on particular migration groups reveals a more interesting picture and is of primary interest. What is most surprising is how clearly the effect of policy seems to be with their intended audience, with the largest effects on foreign-born students. Meanwhile, students with at least one-native born parent seem to have a kind of "buffering" or protective factor in terms of sense of belonging when compared to heritage students.

More specifically, the interaction term for family reunification and fb or nb2fb is slightly negative; perhaps because students in need of such policies have part of their family living abroad and suffer a decreased sense of belonging. Then, both fb and nb2fb students experience increased sense of belonging with policies around education and with anti-discrimination policies in place. Meanwhile, fb students experience slightly decreased levels of sense of belonging with policies around access to permanent

residency, and nationality; both of these policies would only apply to fb students. And there is also a negative relationship with health. While, unsurprising that fb students would experience the largest effect of policy interactions, it is somewhat surprising to see that some policies meant to support migrants (e.g. access to nationality and residency and health) would have a negative relationship, which suggests there may be a disordinal interaction present in the model.[6] Looking closely at model 5, this may be an artifact of the joint/conditional relationship between main- and interactions-effects. Namely, while fb persons have an overall lower sense of belonging compared to heritage students in models 3 & 4, they appear to have a higher level of belonging in model 5. However, when considering the joint-relationship of the regression coefficients that occur simultaneously, they would tend to have lower levels of sense of belonging when compared to heritage students.

Consistent across all groups in the models, regression coefficients are relatively small on ones' overall sense of belonging for policy domains. Yet, in the context of the data, they are of important note. Average policy levels across countries range from 42 – 65 depending on the policy domain and country, as such each policy domain tends to explain about .5 point of the overall belonging for each domain. This is perhaps unsurprising given the small ICC values at the country-level, and is intuitively consistent with expectations that policy is often not directly felt, but is experienced "underneath one's skin." What is most striking is that there is indeed a detectable effect

[6] Likely a disordinal effect would be with the fb group, however the model is too complex to assess simple main effects in this way. The ANOVA parallel would resemble a type of two-way interaction in a fourteen-way ANOVA.

of students, at age 15, of integration policy that is moderated by ones' migration background with foreign-born persons experiencing the greatest effect. In perspective, across the six domains, policies would explain about 3 points of the total belonging scale for fb students, which is about 17%.

Discussion

At present the exploration of immigrant-background as a group warranting measurement is novel in the measurement field, despite a substantial qualitative literature about the unique social and educational needs of "second-generation" migrants and those who emigrate at a young age. Likewise, this work provides an explanatory framework consistent with Zumbo's (2015) ecological model that explicitly recognizes the complexities of our world and improves our ability to make cross-cultural comparisons. By developing a model of multilevel validation of social belonging for peoples of various immigration background, I contribute to a deeper understanding of the Pragmatic view of construct validation (Stone & Zumbo, 2016) and I provide important data in an applied context for policy makers using PISA and other multilevel data.

Fundamental to this project is to understand the way in which social belonging at the individual level informs one's belonging at the country-level. Namely, belonging is often regarded as an individual experience, yet as evidenced by this research, there is also a group component to the experience of belonging in schools that varies by one's migration background; with more recent migrants having lower levels of belonging on average. Moreover, there is a larger socio-political influence on students' sense of

belonging defined by individual country's multicultural policies. This research confirms that increased liberal-democratic policies targeted at immigrant integration are indeed differentially felt based on student's migration background, with more recent immigrants experiencing these policies most; as intended.

Several notes about the practicality of multilevel validation should be discussed. A notable strength of the alignment method is its ability to conduct comparisons across many groups. However, given the iterative nature of estimation using the alignment method, enormous computing power is required. Indeed, with only ten random starts and with 32 groups, the analysis took a full six days of running to converge with unreliable results. Researchers ultimately made use of larger, university supercomputers to conduct the analysis, which still took 2-3 days to complete per model. This was also true for estimation of the three-level models. Similarly, the researchers also attempted to run a measurement model of the six items, across the 12 countries that had English test-taking respondents in a multilevel CFA approach using MPlus 8.2 to ensure no language-effects were impacting the results. Even with a large sample of over 83,000 respondents (including both the full sample and omitting countries with fewer than 1,000 respondents), and both with the four-point response scale and a collapsed binary-response scale; a multilevel model could not converge. In this case, the lack of convergence is primarily attributed to the lack of variance within group. This highlights the tension about how researchers should proceed given known (or suspected) patterns of differential responses, the theories and recommended practices of large-scale data analysis including nested-models, weighting, and

ordinal/categorical data structure, held with the practicalities of real-world data analysis. Yet, in policy analysis and cross-cultural research, we are often dealing a small number of grouping variables, and there is no theoretical larger sample possible without the inclusion of others (e.g. different language groups), which is fraught with additional assumptions.

Indeed, there is an outstanding question about whether or not constructs such as belonging "should" be used at both the individual and societal level given that the construct takes on distinct meanings when moving from one level to the other. However, I take a bit more of a pragmatic approach to the issue. Regardless of whether it should or should not, the use of constructs at multiple levels *is* done, with the OECD being only one such example. Rather, what is incorrect to do is to use evidence from one level for the other. This paper seeks to encourage one to think critically about what the definition of the construct is across level and what claims are appropriate across levels. Then, conceptually, either to embrace the reality of a construct shift as I talk about belonging (and other constructs) and the individual level and then at the country. This is in contrast to other notions, such as economic activity, which is inherently individualistic, but has no shift in construct meaning in the aggregate.

Finally, the novelty of including national-level policy as a part of the ecological framework of construct validation cannot be overstated. Theoretically the field of psychometrics still debates whether the consequences of assessments are a part of the validity debate or rather a part of the validation process (Kane, 2006), but unequivocally there is a recognition that the intended and unintended consequences of assessments

are important (Hubley & Zumbo, 2011; Messick, 1998; Shepard, 1997). Much theoretical lip-service has been offered to say that policies and country-level characteristics are indeed a part of a larger ecological model of item-response and the validation process, but little research has actually been done to demonstrate this connection. This research is perhaps the first of its kind to demonstrate this connection, and it is certainly the first to make the explicit connection between policy and immigrant outcomes in the context of measurement and validity showing that a country's policy context, in part, defines the construct.

Chapter 5: Conclusion- Measurement in Deep Diversity

Collectively, the preceding chapters lay out an argument for the need to think carefully about how we measure and use multi-level constructs as a part of the validation process. In particular, the findings from this research present an empirical example of an ecological framework of item response looking at PISA's measure of social belonging with consideration to students' sense of belonging in school and special attention to one's immigration background.

Review

In Chapter 2, I introduced the concept of social belonging showing how it is an important measure of well-being that predicts a multitude of health and mental health outcomes. Likewise, this chapter connects how perceived group-belonging also, in part, defines one's individual identity. This is particularly true for already marginalized groups like immigrants. As such, immigration background is an important consideration as a social determinant of well-being. In chapter 2, I review the need to create and support social belonging, and the way in which immigration background has historically been assessed. Yet, within the measurement community, it has rarely been considered. Moreover, the consideration of immigration, particularly in large-scale assessment contexts implicitly begins to invoke a multi-level framework. As such, it is important that we adopt methodologies that provide multi-level evidence and that are robust enough in conducting multi-cultural comparison.

In Chapter 3, I introduce a diffractive methodology for comparing multi-method DIF studies and demonstrated that while mostly congruent, the tensions and differences between various methods offer a unique window of analysis to better understand one's construct of interest. In my case, we see that indeed immigration background is an important ecological variable to understand sense of belonging; where item-level analyses using logistic regression suggests some possible methodological artifacts in the scale. Then the scale-level analysis using the alignment method highlights distinct trends across migration background groups, and shows the unique context of "heritage" students compared to those with more complex immigration backgrounds. This investigation also offers additional evidence for careful, but permissible relaxation of the "strict measurement invariance" assumption in order to make cross-cultural measurement comparisons. Namely, in certain circumstances, such as social belonging with respect to immigration background, it is perhaps too strong a criteria to insist on strict measurement invariance. However, it is important to note, that even with a relaxation of the measurement parameters, migration groups *still* cannot be flippantly compared. The use of the alignment method points to the importance of adopting a more thorough multi-level approach with respect to country, which is taken up in Chapter 4.

In Chapter 4, I expand the conversation around validation to argue that claims about multi-level constructs need multi-level evidence, and that country-level policy can, and at times should, be considered a relevant characteristic to understand the ecological context of test respondents. Accounting for cross-cultural language effects,

immigration background, and country test administration, I suggest that it is proper to use a scale-score for PISA's measure of social belonging in school. Then likewise, I introduce specific policy indicators of multi-cultural integration policy that are relevant in the immigration context. As such, I demonstrate that not only do national integration policies have an effect on 15-year-old students in schools, but they have a differential effect based on one's migration background with foreign-born persons experiencing the largest effect of integration policy as intended. Indeed, this may be the first research to introduce policy-level research into the validation process in a cross-cultural context.

Disciplinary Contributions

The consideration of policy levers as introduced in chapter 4, and the recognition of one's socio-political background as an immigrant in chapter 3, together offer a synthesis of the problems and questions introduced in chapter 1. Namely, these empirical studies offer a valuable example of the way in which there can be a beneficial interplay between the fields of political science and measurement. The field of political science, generally speaking, does not seriously engage in psychometric-type questions that are commonplace in the measurement community. Yet, as these studies demonstrate, consideration of immigrant accommodation and acculturation into a broader social fabric necessitate at least some methodological/psychometric consideration in order to confidently investigate socio-political constructs. Likewise, specific policy analyses are typical within political science. However, the need to develop a multi-level evidence base for such a policy analysis is something the

measurement community can readily provide. Moreover, within the field of psychometrics, there is growing demand to seriously consider the importance of response processes (Zumbo & Hubley, 2017). While measurement experts have long recognized how the larger ecological setting, and values around psychological and educational measurement may play a role, there has been little to no empirical demonstration of this. Yet, political scientists have a firm grasp on the way in which country-level policies can be operationalized to target specific groups to (re)shape denizen's experience of their social world; as demonstrated in chapter 4. As such, this dissertation provides an important example of interdisciplinary research that grows both fields simultaneously; not merely using one field in service of the other.

While germane to the field of measurement, validity is a concept that is ultimately relevant for all disciplines. Adopting Kane's (1992) argument based approach, the outstanding question remains, "what is the strength of the validity argument for PISA's measure of social belonging in schools?" For Kane, the argument is constructed in two parts: first is the interpretive argument and second is the validity argument. The interpretive argument puts forward the proposed interpretation and use of the test. In the case of PISA, the interpretative argument is that the PISA measure of social belonging in schools can be used to compare 15-year old students across the globe to help evaluate students' engagement in school. Consistent with my methodology, a higher score indicates a higher sense of belonging on the measure. Used with weights, PISA would make the argument that comparisons can then be made about the national averages across countries. As such, what is the validity argument? My analyses, put

forward here would provide additional evidence to support the presupposed interpretive argument: that PISA *can* compare students' sense of belonging across countries.

However, it is important to make explicit the subtle nuances of this claim, namely PISA does not, at least explicitly, provide an interpretive argument for the use of the belonging scale to compare students with diverse migration backgrounds. Yet, implicitly, this is precisely what is being claimed through the de facto use and publication of these comparisons in PISA. It is this *use* (and others), wherein the validity argument is weak. Indeed, my analysis presented here suggests that *within* some countries, comparisons of migration background groups *should not* be done (see chapter 3). Moreover, if one hopes to compare migration groups across countries, my results (see chapter 4) highlight how the very notion of what it means to belong may take on a distinct meaning for foreign-born and persons with foreign-born parents depending on the degree to which a country adopts and implements liberal democratic policies. One might think of it in a procedural manner, that is to interpret and think about belonging in the colloquial way, immigrants (and children of immigrants), must first gain access to the basic rights of a society (security of residency, health care, education, at least the potential for family unity, etc.) before any equal comparison can be made of them to those without an immigration background. In one sense this is obvious. Yet all too often immigrants are in fact held to a double standard: where they are assessed with reference to the dominant society while not having access, and in many cases actively

denied access, to the basic rights of a democratic society. As such, in these cases, it is my opinion that the validity argument is too weak to make such a comparison.

Canadian Context

While this research is international in nature, it behooves me to make explicit connections of this work to a local context. A multitude of frameworks exists to manage immigration at a national level across the globe, but Canada (and Australia) are particularly unique compared to other countries. Like many developed countries, immigration has always been a part of the Canadian framework to manage population growth and labour. So much so that immigration policy was managed by the Ministry of Mines and Resources from 1936 – 1949. Likewise, much of Canadian immigration had an overt cultural component to it including the *Immigration Act of 1919* that banned certain groups such as communists, Mennonites, Doukhobors, and nationalities that fought against Canada in WWI (Austrians, Hungarians, Turks); the explicit ban on Chinese immigration that formally ended in 1946; the *Immigration Act of 1952* that favoured European and US immigrants. However, in 1967 a new points system was first introduced to rank incoming migrants that is allegedly race, colour, and nationality neutral; but language, education, and past work experience are considered; which of course are partially shaped by race, colour, and nationality. Under former Prime Minister Pierre Trudeau, the *Immigration Act of 1976* provided an even further break from the past by continuing the economic, social, and cultural goals but also prioritizing family reunification, diversity, recognition of refugees as a distinct class, and integrated

non-discrimination. These changes, and the inclusion of a semi-independent immigration judicial board, help to define a radically new platform for immigration in comparison to other developed countries. And it is this framework that Canada continues to seek to legitimate. Between 1996 and 2016, two-thirds of Canadian population growth is attributable to immigration. Unquestionably, Canada has succeeded in their ability to attract and bring immigrants into the country. Now, an important questions remain to test this framework: how can Canada retain immigrants? Maintain or increase the social fabric of all denizens? Do immigrants feel that they belong? This is especially important in light of the "type" of immigrants Canada is welcoming. "Economic Class" immigrants represented well over half of all incoming immigrants in 2016.

By nature of the Canadian selection process, these immigrants are some of the most educated and most skilled people, and accordingly they are some of the most mobile. Meaning, should they and/or their family no longer feel a sense of belonging to the Canadian political project, we should not assume they will remain in Canada, since they have already demonstrated their capacity to emigrate under one of the most scrutinized immigration regimes. This highlights the complex relationship between government policy and individual desire. As demonstrated in Chapter 4, government policy has a notable effect on individual sense of belonging, especially for recent arrivals. Then, at least on a conceptual basis, there is a feedback loop where immigrants' sense of belonging also informs the creation of government policy in an attempt to retain immigrants in Canada. As such, a better understanding of sense of belonging,

especially for recent arrivals, is important to evaluate the Canadian immigration framework.

Education Theory

This discussion of the Canadian context is relevant as one particular example to think about the construct of social belongingness broadly speaking; as discussed in chapter 2. In chapter 2, I explored past research showing the importance of social belonging and its relevance as a social determinant for both individual and group-level outcomes. This research highlights how the complexity of social belongness as a sociological construct and demonstrates the multi-level nature of the construct. In particular, thorough investigation of social belongingness, particularly in a large-scale context, benefits from consideration of one's immigration background. Likewise, in considering the ecologically setting, country-level characteristics are valuable to consider, with multicultural, integration policy being one such example. But what do these findings mean for a deeper understanding of social belongingness altogether and what are the implications in a methodological sense? One thing that has emerged in this research is the utility of framing an ecological model as a subset within a framework of situative theory. The evidence presented here demonstrates the extent to which it behooves the field of measurement to adopt and integrate "situative theory" in our approach to the assessment of psychological constructs. As put forth by Vygotsky (1978) and developed by Nasir and Saxe (2003), Anderson et. al (2000), and others, our learning and meaning making is inherently a social process. Indeed, while not self-prescribed, one could read Stone and Zumbo (2016) as engaging in a framework of

situative theory. Nowhere is this perhaps more germane than when discussing immigration. Immigrants are, by definition, people coming from a different social background that is already highly contextualized based on what their country of origin might be and what their "in-group" belonging might be in their country of origin and then their host community. Furthermore, the construct of social belonging also necessitates a consideration for the varied and complex social ties that define the social fabric. What this means for the measurement community, is that we must work harder to develop theories of measurement, tools of measurement, and methodologies for analyses that are flexible, adaptable, and sensitive enough to the "situated" position of the measurement processes being captured. This investigation puts forward only a very beginning and small step in this direction. By further unpacking the ecologically setting of the measurement of social belonging including individual and group-level characteristics related to immigration, and by capturing country-level characteristics through integration policy, I aim to better situate our understanding of social belonging as a process. By advancing the use of an "approximate" measurement approach that tolerates some measurement non-invariance and building on multi-level measurement evidence, I seek to show how existing tools can be appropriated to support a richer understanding of psychological measurement. Yet, much more needs to be done.

Limitations and Future Directions

How do we better capture the complexity of individuals in large-scale data analysis? Perhaps through novel ways of tolerating small group sample sizes? Or perhaps renewed inquiry into Bayesian statistical approaches can better support our

inferences that do not rely on our frequentist assumptions of population-level inferences? All of which come with substantial implications, one of which includes our computing capacity; which I have previously discussed. It is these types of methodological considerations that I seek to develop in future work. Then likewise, within the field of migration, the new (use of) tools will better support our ability to understand the sociological fabric which is being co-created by our diverse denizens.

Indeed, organizations and researchers that collect and/or use large-scale assessment data like the OECD have a responsibility to define the full scope of inferences that can and cannot be made. In the case of this particular scale, the OECD can and does test the cross-country measurement invariance to some extent with evidence that suggests that scale can be used to make inferences. However, more detailed analysis, like those performed in this dissertation, highlight the limits of inference making using a subgroup analyses, like migration background. Furthermore, there is a trade-off in research design about modifying scales and assessments given the desire to maintain comparability across time with past administrations of PISA or other tools. However, as this analysis and others (e.g. Y. Chen, 2017) show, new evidence about scale development has shed light on ways to reduce construct-irrelevant variance in the data, such as avoiding negatively framed items. Modifications of this kind are relatively simple and would improve the psychometric properties for the foreseeable future.

Future analyses would benefit by extending the scope of variables and their relationships. It is difficult to get variables that are comparable enough across countries

and their respective contexts (e.g. the definition of "immigrant" is different in Korea vs. Canada), which is why international datasets like the OECD are useful by creating equivalence. Further analysis of the context of "immigrants", such as refugee status, age of arrival, and duration of stay in the host country, would all tremendously improve the scope of nuance of this analysis and others. Likewise, variables to understand the degree of ethnic diversity at the school-level and to what extent one's background matches the dominant minority group might also shed light on social belonging.

Likewise, in consideration of the diffractive methodology put forward here, future work needs to be done to explore a formal, albeit necessarily adaptive, framework in the use of multiple methodologies. How can we develop sophisticated ways to distinguish between noise and error and methodological tensions that are themselves worthy of consideration? And in particular how do we better employ individualized, situated knowledge within large scale data analysis? Can ethnographic work be used as a kind of "prior" in a Bayesian approach? And if so, how do we integrate these perspectives? It is my belief that we can, as psychometricians and methodologist, advance a new perspective on the field of measurement that is creative enough to handle this complexity. Yet, likely, to paraphrase the religious philosopher C. S. Lewis, if we are on the wrong road progress means doing an about-turn and walking back to the right road. In that case, the methodologist who turns back soonest achieves the most progress. In our case, this means revisiting some of our assumptions of statistical science and absolutely means we must unlearn some of our "positivist" tendencies that are latent within our field. A diffractive approach to data science and

psychometrics is about the art of argumentation and evidence building. It will require a flexibility to "read between the lines" of various methodological approaches. And it will probably be readily rejected as an affront to established norms of psychometrics.

Novel Contributions

It is perhaps this last point that is the biggest contribution of this dissertation. As a review, throughout the research, I have made several significant contributions to the fields of psychometrics and political science by demonstrating the benefits from greater integration of these two fields. On the more technical side, I have provided an extension of the alignment method with polytomous-type data, and I have developed a multi-level framework of policy analysis as a form of construct validation; which is certainly one of the first of its kind. I have explored the limits of our statistical computing capacity, and I have integrated multiple methods into a complex story about sense of belonging with respect to immigration background. I have put forward empirical examples about the importance of an enriched, ecological framework that includes policy characteristics of countries to be used in large-scale data assessment, and demonstrated an initial way to start such an investigation and I have suggested a way to think about statistics through a lens of situative theory. But more subtly, throughout the dissertation I am engaging in philosophical questions about the role of statistical inference and the ability to adopt a situative framework that better respects individual lived experiences using the example of immigration and social belonging. Through this I hope to advance a more philosophical type of argumentation consistent with the scope of a "doctor of philosophy." It is my belief, grounded in the empirical study presented

here, my professional experience, my education as a student at the University of British Columbia and other institutions, and as a visitor and learner on the unceded territories of the Musqueam peoples; that we are living in an age that requires unprecedented levels of grace and creativity to navigate through the "noise" of statistics and data to contribute to knowledge creation. I believe that knowledge is indeed place-based, and that localized ways of knowing and meaning making can and do have a role to play in large-scale data analysis that informs our decision making for the collective. It is imperative that we develop the ability to make inferences that are sound and that respect the larger ecological setting of learning and meaning-making; this is how we do measurement in an age of deep diversity.

References

Adam, F. (2008). Mapping social capital across Europe: Findings, trends and methodological shortcomings of cross-national surveys. *Social Science Information*, *47*(2), 159–186. https://doi.org/10.1177/0539018408089077

Alba, R., & Foner, N. (2014). Comparing Immigrant Integration in North America and Western Europe: How Much Do the Grand Narratives Tell Us? *International Migration Review*, *48*, S263–S291. https://doi.org/10.1111/imre.12134

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.) (Eds.). (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Anderson, J. R., Greeno, J. G., Reder, L. M., & Simon, H. A. (2000). Perspectives on Learning, Thinking, and Activity. *Educational Researcher*, *29*(4), 11–13. https://doi.org/10.3102/0013189X029004011

Ariely, G., & Davidov, E. (2011). Can we Rate Public Support for Democracy in a Comparable Way? Cross-National Equivalence of Democratic Attitudes in the World Value Survey. *Social Indicators Research*, *104*(2), 271–286.

Aron, A., Aron, E. N., Tudor, M., & Nelson, G. (1991). Close relationships as including other in the self. *Journal of Personality and Social Psychology*, *60*(2), 241–253.

Asparouhov, T., & Muthén, B. (2014). Multiple-Group Factor Analysis Alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(4), 495–508. https://doi.org/10.1080/10705511.2014.919210

Banting, K., & Kymlicka, W. (2013). Is there really a retreat from multiculturalism policies? New evidence from the multiculturalism policy index. *Comparative European Politics; Basingstoke*, *11*(5), 577–598. http://dx.doi.org/10.1057/cep.2013.12

Barad, K. (2007). *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. https://doi.org/10.1215/9780822388128

Baubock, R. (2003). A Closer Look at Unity: What Social Cohesion Does and Does Not Require. *Responsive Community*, *13*(2), 31–42.

Baumeister, R. F. (1992). *Meanings of Life* (1 edition). New York: The Guilford Press.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*(3), 497–529. https://doi.org/10.1037/0033-2909.117.3.497

Baumeister, R. F., Twenge, J. M., & Nuss, C. K. (2002). Effects of social exclusion on cognitive processes: Anticipated aloneness reduces intelligent thought. *Journal of Personality and Social Psychology*, *83*(4), 817–827. https://doi.org/10.1037/0022-3514.83.4.817

Beauducel, A., & Herzberg, P. Y. (2006). On the Performance of Maximum Likelihood Versus Means and Variance Adjusted Weighted Least Squares Estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(2), 186–203. https://doi.org/10.1207/s15328007sem1302_2

Benet-Martínez, V., Lee, F., & Leu, J. (2006). Biculturalism and Cognitive Complexity. *Journal of Cross-Cultural Psychology*, *37*(4), 386–407. https://doi.org/10.1177/0022022106288476

Berkman, L. F., & Syme, S. L. (1979). Social Networks, Host Resistance, and Mortality: A Nine-Year Follow-Up Study of Alameda County Residents. *American Journal of Epidemiology*, *109*(2), 186–204. https://doi.org/10.1093/oxfordjournals.aje.a112674

Bernhardt, P. C., Dabbs Jr, J. M., Fielden, J. A., & Lutter, C. D. (1998). Testosterone changes during vicarious experiences of winning and losing among fans at sporting events. *Physiology & Behavior*, *65*(1), 59–62. https://doi.org/10.1016/S0031-9384(98)00147-4

Berry, J. W. (2005). Acculturation: Living successfully in two cultures. *International Journal of Intercultural Relations*, *29*(6), 697–712. https://doi.org/10.1016/j.ijintrel.2005.07.013

Bertossi, C., & Duyvendak, J. W. (2012). National models of immigrant integration: The costs for comparative research. *Comparative European Politics*, *10*(3), 237–247. https://doi.org/10.1057/cep.2012.10

Bjørnskov, C. (2007). Determinants of Generalized Trust: A Cross-Country Comparison. *Public Choice*, *130*(1/2), 1–21.

Bjørnskov, C. (2008). Social Trust and Fractionalization: A Possible Reinterpretation. *European Sociological Review*, *24*(3), 271–283.

Blank, R. M. (2005). Tracing the Economic Impact of Cumulative Discrimination. *The American Economic Review*, *95*(2), 99–103.

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations.* (pp. 349–381). San Francisco: Jossey-Bass.

Bloemraad, I. (2006). *Becoming a Citizen: Incorporating Immigrants and Refugees in the United States and Canada* (1 edition). Berkeley: University of California Press.

Bloemraad, I. (2015). Theorizing and Analyzing Citizenship in Multicultural Societies. *The Sociological Quarterly*, *56*(4), 591–606. https://doi.org/10.1111/tsq.12095

Bloemraad, I., Korteweg, A., & Yurdakul, G. (2008). Citizenship and Immigration: Multiculturalism, Assimilation, and Challenges to the Nation-State. *Annual Review of Sociology*, *34*(1), 153–179. https://doi.org/10.1146/annurev.soc.34.040507.134608

Bloemraad, I., & Wright, M. (2014). "Utter Failure" or Unity out of Diversity? Debating and Evaluating Policies of Multiculturalism. *International Migration Review*, *48*, S292–S334. https://doi.org/10.1111/imre.12135

Blum, R. W., & Libbey, H. P. (2004). Executive Summary. *Journal of School Health*, *74*(7), 231–232. https://doi.org/10.1111/j.1746-1561.2004.tb08278.x

Bollen, K. A. (1989). *Structural Equations with Latent Variables* (1 edition). New York: Wiley-Interscience.

Breen, R., & Jonsson, J. O. (2005). Inequality of Opportunity in Comparative Perspective: Recent Research on Educational Attainment and Social Mobility. *Annual Review of Sociology*, *31*(1), 223–243. https://doi.org/10.1146/annurev.soc.31.041304.122232

Bronfenbrenner, U. (1981). *The Ecology of Human Development: Experiments by Nature and Design* (1981 edition). Cambridge, Mass: Harvard University Press.

Brubaker, R. (1998). *Citizenship and Nationhood in France and Germany*. Cambridge, Mass.: Harvard University Press.

Byrne, B. M., & van de Vijve, F. J. R. (2017). The maximum likelihood alignment approach to testing for approximate measurement invariance: A paradigmatic cross-cultural application. *Psicothema*, *29*(4). Retrieved from http://www.redalyc.org/resumen.oa?id=72753218017

Cacioppo, J. T., & Patrick, W. (2009). *Loneliness: Human Nature and the Need for Social Connection* (Reprint edition). New York London: W. W. Norton & Company.

Chen, F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chen, F., & West, S. G. (2008). Measuring individualism and collectivism: The importance of considering differential components, reference groups, and measurement invariance. *Journal of Research in Personality*, *42*(2), 259–294. https://doi.org/10.1016/j.jrp.2007.05.006

Chen, G., Matthieu, J. E., & Bliese, P. D. (2004). A framework for conducting multilevel construct validation. In F. J. Yammarino & F. Dansereau (Series Ed.), *Research in multilevel issues: Multilevel issues in organizational behavior and processes* (Vol. 3, pp. 273–303). Oxford, UK: Elsevier.

Chen, Y. (2017). *On the impact of negatively keyed items on the assessment of the unidimensionality of psychological tests and measures* (University of British Columbia). https://doi.org/10.14288/1.0357244

Cherry, A. L. (1994). *The Socializing Instincts: Individual, Family, and Social Bonds*. Westport, Conn: Praeger.

Chiu, M. M., Pong, S., Mori, I., & Chow, B. W. (2012). Immigrant Students' Emotional and Cognitive Engagement at School: A Multilevel Analysis of Students in 41 countries. *Journal of Youth and Adolescence; New York*, *41*(11), 1409–1425. http://dx.doi.org.ezproxy.library.ubc.ca/10.1007/s10964-012-9763-x

Cohen, G. L., & Garcia, J. (2005). “I Am Us”: Negative Stereotypes as Collective Threats. *Journal of Personality and Social Psychology*, *89*(4), 566–582. https://doi.org/10.1037/0022-3514.89.4.566

Cohen, G. L., Garcia, J., Purdie-Vaughns, V., Apfel, N., & Brzustoski, P. (2009). Recursive Processes in Self-Affirmation: Intervening to Close the Minority Achievement Gap. *Science*, *324*(5925), 400–403. https://doi.org/10.1126/science.1170769

Cohen, G. L., Steele, C. M., & Ross, L. D. (1999). The Mentor’s Dilemma: Providing Critical Feedback Across the Racial Divide. *Personality and Social Psychology Bulletin*, *25*(10), 1302–1318. https://doi.org/10.1177/0146167299258011

Cohen, S., & Janicki-Deverts, D. (2009). Can We Improve Our Physical Health by Altering Our Social Networks? *Perspectives on Psychological Science*, *4*(4), 375–378.

Coleman, J. S. (1988). Social Capital in the Creation of Human Capital. *American Journal of Sociology*, *94*, S95–S120.

Coon, C. S. (1946). The Universality of Natural Groupings in Human Societies. *The Journal of Educational Sociology*, *20*(3), 163–168. https://doi.org/10.2307/2263780

Council of Europe. (2008). *White Paper on Intercultural Dialogue: Living Together as Equals in Dignity*. Retrieved from http://www.coe.int/t/dg4/intercultural/publication_whitepaper_id_EN.asp

Council of the European Union. (2004). *Immigrant Integration Policy in the European Union*. Retrieved from http://data.consilium.europa.eu/doc/document/ST-14615-2004-INIT/en/pdf

Cronbach, L. J. (1988). Five perspectives on validity argument. In H Wainer & H. Braun (Eds.), *Test Validity* (pp. 3–17). Hillsdale, NJ: Lawrence Erlbaum.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

Da Costa, P. D., & Araújo, L. (2012). Differential item functioning (DIF): What function differently for Immigrant students in PISA 2009 reading items. *JRC Scientific and Policy Reports. Luxembourg: European Commission*.

Dansereau, F., Alutto, J. A., & Yammarino, F. J. (1984). *Theory testing in organizational behavior: The variant approach.* Englewood Cliffs, NJ.: Prentice Hall.

Dansereau, F., & Yammarino, F. J. (2000). Within and between analysis: The variant paradigm as an underlying approach to theory building and testing. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 425–466).

Davidov, E. (2009). Measurement Equivalence of Nationalism and Constructive Patriotism in the ISSP: 34 Countries in a Comparative Perspective. *Political Analysis*, *17*(1), 64–82.

Davidov, E., Cieciuch, J., Meuleman, B., Schmidt, P., Algesheimer, R., & Hausherr, M. (2015). The Comparability of Measurements of Attitudes toward Immigration in the European Social Survey Exact versus Approximate Measurement Equivalence. *Public Opinion Quarterly*, *79*(S1), 244–266. https://doi.org/10.1093/poq/nfv008

Dixon-Román, E. J. (2017). *Inheriting Possibility: Social Reproduction and Quantification in Education* (1 edition). Minneapolis: Univ Of Minnesota Press.

Dworkin, R. (1985). *A Matter of Principle*. Harvard University Press.

Eccles, J. S., & Roeser, R. W. (2011). Schools as Developmental Contexts During Adolescence. *Journal of Research on Adolescence*, *21*(1), 225–241. https://doi.org/10.1111/j.1532-7795.2010.00725.x

Entzinger, H. (2014). The growing gap between facts and discourse on immigrant integration in the Netherlands. *Identities*, *21*(6), 693–707. https://doi.org/10.1080/1070289X.2013.828616

Finn, J. D. (1989). Withdrawing From School. *Review of Educational Research*, *59*(2), 117–142. https://doi.org/10.3102/00346543059002117

Flake, J. K., & McCoach, D. B. (2018). An Investigation of the Alignment Method With Polytomous Indicators Under Conditions of Partial Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(1), 56–70. https://doi.org/10.1080/10705511.2017.1374187

Forer, B., & Zumbo, B. D. (2011). Validation of Multilevel Constructs: Validation Methods and Empirical Findings for the EDI. *Social Indicators Research*, *103*(2), 231. https://doi.org/10.1007/s11205-011-9844-3

Fox, J. D. (2003). From Products to Process: An Ecological Approach to Bias Detection. *International Journal of Testing*, *3*(1), 21–47. https://doi.org/10.1207/S15327574IJT0301_2

Freeman, G. P. (2004). Immigrant Incorporation in Western Democracies. *The International Migration Review*, *38*(3), 945–969.

Freitag, M., & Bauer, P. C. (2013). Testing for Measurement Equivalence in Surveys: Dimensions of Social Trust Across Cultural Contexts. *Public Opinion Quarterly*, *77*(S1), 24–44. https://doi.org/10.1093/poq/nfs064

Gadermann, A. M., Guhn, M., Schonert-Reichl, K. A., Hymel, S., Thomson, K., & Hertzman, C. (2016). A Population-Based Study of Children's Well-Being and Health: The Relative

Importance of Social Relationships, Health-Related Activities, and Income. *Journal of Happiness Studies*, *17*(5), 1847–1872. https://doi.org/10.1007/s10902-015-9673-1

Goffman, E. (1986). *Stigma: Notes on the Management of Spoiled Identity* (Reissue edition). New York: Touchstone.

Gómez-Benito, J., Sireci, S., & Padilla, J.-L. (2018). Differential Item Functioning: Beyond validity evidence based on internal structure. *Psicothema*, (30.1), 104–109. https://doi.org/10.7334/psicothema2017.183

Goodenow, C. (1993). The psychological sense of school membership among adolescents: Scale development and educational correlates. *Psychology in the Schools*, *30*(1), 79–90.

Goodman, S. W. (2010). Integration Requirements for Integration's Sake? Identifying, Categorising and Comparing Civic Integration Policies. *Journal of Ethnic and Migration Studies*, *36*(5), 753–772. https://doi.org/10.1080/13691831003764300

Guhn, M., Schonert-Reichl, K. A., Gadermann, A. M., Marriott, D., Pedrini, L., Hymel, S., & Hertzman, C. (2012). Well-Being in Middle Childhood: An Assets-Based Population-Level Research-to-Action Project. *Child Indicators Research*, *5*(2), 393–418. https://doi.org/10.1007/s12187-012-9136-8

Hagerty, B. M., Lynch-Sauer, J., Patusky, K. L., Bouwsema, M., & Collier, P. (1992). Sense of belonging: A vital mental health concept. *Archives of Psychiatric Nursing*, *6*(3), 172–177.

Hamachek, D. E. (1987). *Encounters with the Self* (3rd Revised edition). New York: Thomson Learning.

Hamre, B. K., & Pianta, R. C. (2005). Can Instructional and Emotional Support in the First-Grade Classroom Make a Difference for Children at Risk of School Failure? *Child Development*, *76*(5), 949–967.

Harell, A., & Stolle, D. (2010). Diversity and Democratic Politics: An Introduction. *Canadian Journal of Political Science / Revue Canadienne de Science Politique*, *43*(2), 235–256.

Helbling, M. (2013). Validating integration and citizenship policy indices. *Comparative European Politics*, *11*(5), 555–576. https://doi.org/10.1057/cep.2013.11

Hertzman, C., & Boyce, T. (2010). How experience gets under the skin to create gradients in developmental health. *Annual Review of Public Health*, *31*, 329–347. https://doi.org/10.1146/annurev.publhealth.012809.103538

Hoffman, M., Richmond, J., Morrow, J., & Salomone, K. (2002). Investigating "Sense of Belonging" in First-Year College Students. *Journal of College Student Retention: Research, Theory & Practice*, *4*(3), 227–256. https://doi.org/10.2190/DRYC-CXQ9-JQ8V-HT4V

Holland, P. W., & Thayer, D. T. (1988). Differential Item Functioning and the Mantel-Haenszel Procedure. In Howard Wainer & H. I. Braun (Eds.), *Test Validity* (pp. 129–145). Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/j.2330-8516.1986.tb00186.x/abstract

Hooghe, M., Reeskens, T., Stolle, D., & Trappers, A. (2009). Ethnic Diversity and Generalized Trust in Europe: A Cross-National Multilevel Study. *Comparative Political Studies*, *42*(2), 198–223. https://doi.org/10.1177/0010414008325286

Howie, S. J., & Plomp, T. (2006). Lessons from cross-national research on context and achievement: Hunting and fishing in the TIMSS landscape. In S. J. Howie & T. Plomp

(Eds.), *Contexts of Learning Mathematics and Science: Lessons Learned from TIMSS* (pp. 3–15). New York, NY: Routledge.

Hox, J. J., Moerbeek, M., & Schoot, R. van de. (2017). *Multilevel Analysis: Techniques and Applications, Third Edition* (3 edition). Routledge.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hubley, A. M., & Zumbo, B. D. (2011). Validity and the Consequences of Test Interpretation and Use. *Social Indicators Research*, *103*(2), 219–230.

Hummer, R. A., Rogers, R. G., Nam, C. B., & LeClere, F. B. (1999). Race/Ethnicity, Nativity, and U.S. Adult Mortality. *Social Science Quarterly*, *80*(1), 136–153.

IOM. (2018, October 30). Global Migration Trends. Retrieved June 20, 2019, from International Organization for Migration website: https://www.iom.int/global-migration-trends

IRCC, I., Refugees and Citizenship Canada. (2017, November 1). Notice – Supplementary Information 2018-2020 Immigration Levels Plan [Notices]. Retrieved February 3, 2018, from Aem website: https://www.canada.ca/en/immigration-refugees-citizenship/news/notices/supplementary-immigration-levels-2018.html

Jethwani-Keyser, M. M. (2008). *"When teachers treat me well, I think I belong": School belonging and the psychological and academic well-being of adolescent girls in urban India* (Doctoral Dissertation). New York University.

Jilke, S., Meuleman, B., & Van de Walle, S. (2015). We Need to Compare, but How? Measurement Equivalence in Comparative Public Administration. *Public Administration Review*, *75*(1), 36–48. https://doi.org/10.1111/puar.12318

Johnson, D. R., Alvarez, P., Longerbeam, S., Soldner, M., Inkelas, K. K., Leonard, J. B., & Rowan-Kenyon, H. (2007). Examining sense of belonging among first-year undergraduates from different racial/ethnic groups. *Journal of College Student Development*, *48*(5), 525–542.

Joppke, C. (2007). Beyond national models: Civic integration policies for immigrants in Western Europe. *West European Politics*, *30*(1), 1–22. https://doi.org/10.1080/01402380601019613

Joppke, C. (2010). *Citizenship and Immigration* (1 edition). Cambridge; Malden, MA: Polity.

Kane, M. (1992). An argument-based approach to validity. *Psychological Bulletin*, *112*(3), 527–535. https://doi.org/10.1037/0033-2909.112.3.527

Kane, M. (2006). Validation. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: American Council on Education and Praeger.

Kankaraš, M., & Moors, G. (2014). Analysis of Cross-Cultural Comparability of PISA 2009 Scores. *Journal of Cross-Cultural Psychology*, *45*(3), 381–399. https://doi.org/10.1177/0022022113511297

Kankaraš, M., Vermunt, J. K., & Moors, G. (2011). Measurement Equivalence of Ordinal Items: A Comparison of Factor Analytic, Item Response Theory, and Latent Class Approaches. *Sociological Methods & Research*, *40*(2), 279–310. https://doi.org/10.1177/0049124111405301

Kesler, C., & Bloemraad, I. (2010). Does Immigration Erode Social Capital? The Conditional Effects of Immigration-Generated Diversity on Trust, Membership, and Participation across 19 Countries, 1981-2000. *Canadian Journal of Political Science / Revue Canadienne de Science Politique*, *43*(2), 319–347.

Kiecolt-Glaser, J. K., Fisher, L. D., Ogrocki, P., Stout, J. C., Speicher, C. E., & Glaser, R. (1987). Marital quality, marital disruption, and immune function. *Psychosomatic Medicine*, *49*(1), 13–34.

Klein, K. J., Dansereau, F., & Hall, R. J. (1994). Levels Issues in Theory Development, Data Collection, and Analysis. *The Academy of Management Review*, *19*(2), 195–229. https://doi.org/10.2307/258703

Koopmans, R. (2010). Trade-Offs between Equality and Difference: Immigrant Integration, Multiculturalism and the Welfare State in Cross-National Perspective. *Journal of Ethnic and Migration Studies*, *36*(1), 1–26. https://doi.org/10.1080/13691830903250881

Koopmans, R., Statham, P., Giugni, M., & Passy, F. (2005). *Contested Citizenship: Immigration and Cultural Diversity in Europe*. Retrieved from http://muse.jhu.edu/book/31718

Lee, R. M., & Robbins, S. B. (1995). Measuring belongingness: The Social Connectedness and the Social Assurance scales. *Journal of Counseling Psychology*, *42*(2), 232–241. https://doi.org/10.1037/0022-0167.42.2.232

Lee, R. M., & Robbins, S. B. (1998). The relationship between social connectedness and anxiety, self-esteem, and social identity. *Journal of Counseling Psychology*, *45*(3), 338–345. https://doi.org/10.1037/0022-0167.45.3.338

Lissitz, R. W. (2009). *The concept of validity*. Charlotte, NC: Information Age Publishing Inc.

Lyubomirsky, S., Sheldon, K. M., & Schkade, D. (2005). Pursuing happiness: The architecture of sustainable change. *Review of General Psychology*, *9*(2), 111–131. https://doi.org/10.1037/1089-2680.9.2.111

Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., Muthén, B., & Dicke, T. (2018). What to do when scalar invariance fails: The extended alignment method for

multi-group factor analysis comparison of latent means across many groups. *Psychological Methods*, *23*(3), 524–545. https://doi.org/10.1037/met0000113

Maxwell, R. (2010). Evaluating Migrant Integration: Political Attitudes Across Generations in Europe. *The International Migration Review*, *44*(1), 25–52.

Mendoza-Denton, R., Downey, G., Purdie, V. J., Davis, A., & Pietrzak, J. (2002). Sensitivity to status-based rejection: Implications for African American students' college experience. *Journal of Personality and Social Psychology*, *83*(4), 896–918. https://doi.org/10.1037/0022-3514.83.4.896

Messick, S. (1998). Test validity: A matter of consequences. *Social Indicators Research*, *45*, 35–44.

Mok, A., Morris, M. W., Benet-Martínez, V., & Karakitapoğlu-Aygün, Z. (2007). Embracing American Culture: Structures of Social Identity and Social Networks Among First-Generation Biculturals. *Journal of Cross-Cultural Psychology*, *38*(5), 629–635. https://doi.org/10.1177/0022022107305243

Morgenthau, H. J. (1962, March 1). Love and Power. Retrieved July 29, 2017, from Commentary Magazine website: https://www.commentarymagazine.com/articles/love-and-power/

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*(3), 313–335. https://doi.org/10.1037/a0026802

Muthén, L. (2018, May 24). *Model identification*.

Nasir, N. S. (2008). Everyday Pedagogy: Lessons from Basketball, Track, and Dominoes. *Phi Delta Kappan*, *89*(7), 529–532. https://doi.org/10.1177/003172170808900717

Nasir, N. S., & Saxe, G. B. (2003). Ethnic and Academic Identities: A Cultural Practice Perspective on Emerging Tensions and Their Management in the Lives of Minority Students. *Educational Researcher*, *32*(5), 14–18.

Nee, R., & Alba, V. (2003). *Remaking the American Mainstream: Assimilation and Contemporary Immigration*. Harvard University Press.

Newcomb, M. D. (1990). Social Support by Many Other Names: Towards a Unified Conceptualization. *Journal of Social and Personal Relationships*, *7*(4), 479–494. https://doi.org/10.1177/0265407590074005

Nguyen, A.-M. D., & Benet-Martínez, V. (2013). Biculturalism and Adjustment: A Meta-Analysis. *Journal of Cross-Cultural Psychology*, *44*(1), 122–159. https://doi.org/10.1177/0022022111435097

Nuñez, A.-M. (2009). A critical paradox? Predictors of latino students' sense of belonging in college. *Journal of Diversity in Higher Education*, *2*(1), 46–61. https://doi.org/10.1037/a0014099

Oberle, E., Schonert-Reichl, K. A., Guhn, M., Zumbo, B. D., & Hertzman, C. (2014). The Role of Supportive Adults in Promoting Positive Development in Middle Childhood: A Population-Based Study. *Canadian Journal of School Psychology; Vancouver*, *29*(4), 296–316.

O'Connor, B. P. (2019). Within-and-Between-Analysis for Groups and Dyads (Version 1.0) [R].

OECD. (2017a). *Education at a Glance 2017: OECD Indicators*. Retrieved from http://www.oecd.org/education/education-at-a-glance-19991487.htm

OECD. (2017b). *PISA 2015 Results (Volume III): Students' Well-Being*. Retrieved from http://dx.doi.org/10.1787/9789264273856-en

Oliveri, M. E., & Ercikan, K. (2011). Do Different Approaches to Examining Construct Comparability in Multilanguage Assessments Lead to Similar Conclusions? *Applied Measurement in Education*, *24*(4), 349–366. https://doi.org/10.1080/08957347.2011.607063

Oliveri, M. E., Olson, B. F., Ercikan, K., & Zumbo, B. D. (2012). Methodologies for Investigating Item- and Test-Level Measurement Equivalence in International Large-Scale Assessments. *International Journal of Testing*, *12*(3), 203–223. https://doi.org/10.1080/15305058.2011.617475

Orbell, J., van de Kragt, A., & Dawes, R. (1988). Explaining Discussion-Induced Cooperation. *Journal of Personality and Social Psychology*, *54*, 811–819.

Parekh. (2006). *Rethinking Multiculturalism*. Palgrave Macmilan, Paperback.

Patton, M. J., & And Others. (1982). Kohut's Psychology of the Self: Theory and Measures of Counseling Outcome. *Journal of Counseling Psychology*, *29*(3), 268–282.

Pinel, E. C. (1999). Stigma consciousness: The psychological legacy of social stereotypes. *Journal of Personality and Social Psychology*, *76*(1), 114–128. https://doi.org/10.1037/0022-3514.76.1.114

Portes, A., & Rumbaut, R. G. (2001). *Legacies: The Story of the Immigrant Second Generation* (First Printing edition). Berkeley: New York: University of California Press.

Portes, A., & Vickstrom, E. (2011). Diversity, Social Capital, and Cohesion. *Annual Review of Sociology*, *37*, 461–479.

powell, john a., & Menendian, S. (2006). The Problem of Othering: Towards Inclusiveness and Belonging. *Othering & Belonging: Expanding the Circle of Human Concern*, *1*(1), 16–39.

Putnam, R. D. (2007). E Pluribus Unum: Diversity and Community in the Twenty-first Century The 2006 Johan Skytte Prize Lecture. *Scandinavian Political Studies*, *30*(2), 137–174. https://doi.org/10.1111/j.1467-9477.2007.00176.x

Redhead, M. (2002). *Charles Taylor: Thinking and Living Deep Diversity*. Lanham, Md: Rowman & Littlefield Publishers.

Reis, H. T., & Franks, P. (1994). The role of intimacy and social support in health outcomes: Two processes or one? *Personal Relationships*, *1*(2), 185–197. https://doi.org/10.1111/j.1475-6811.1994.tb00061.x

Roberson, N. D., & Zumbo, B. D. (2019). Migration Background in PISA's Measure of Social Belonging: Using a Diffractive Lens to Interpret Multi-Method DIF Studies. *International Journal of Testing*, *0*(0), 1–26. https://doi.org/10.1080/15305058.2019.1632316

Rucht, D., & Neidhardt, F. (2001). Protestgeschichte der Bundesrepublik Deutschland, 1950-1994: Ereignisse, Themen, Akteure. In D. Rucht (Ed.), *Protest in der Bundesrepublik: Strukturen und Entwicklungen* (pp. 27–70). Frankfurt am Main: Campus Verlag.

Sandercock, L. (2003). *Cosmopolis II: Mongrel Cities of the 21st Century* (2nd edition). New York: Bloomsbury Academic.

Shepard, L. A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice*, *16*(2), 5–24.

Sherif, M., Harvey, O. J., Hood, W. R., Sherif, C. W., & White, J. (1988). *The Robbers Cave Experiment: Intergroup Conflict and Cooperation.* Middletown, Conn. : Scranton, Pa: Wesleyan.

Slocum-Gori, S. L., & Zumbo, B. D. (2011). Assessing the Unidimensionality of Psychological Scales: Using Multiple Criteria from Factor Analysis. *Social Indicators Research*, *102*(3), 443–461.

Solomon, Z., Waysman, M., & Mikulincer, M. (1990). Family Functioning, Perceived Societal Support, and Combat-Related Psychopathology: The Moderating Role of Loneliness. *Journal of Social and Clinical Psychology*, *9*(4), 456–472. https://doi.org/10.1521/jscp.1990.9.4.456

Spaulding, J. A., Simpson, G., & Durkheim, E. (1997). *Suicide* (1 edition). New York: Free Press.

Spencer, S., & Cooper, B. (2006). *Social Integration of Migrants in Europe: A Review of the European Literature*. Retrieved from https://www.compas.ox.ac.uk/media/ER-2006-Integration_Europe_Literature_Review_OECD.pdf

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613–629. https://doi.org/10.1037/0003-066X.52.6.613

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797–811. https://doi.org/10.1037/0022-3514.69.5.797

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing Measurement Invariance in Cross-National Consumer Research. *Journal of Consumer Research*, *25*(1), 78–90. https://doi.org/10.1086/209528

Stone, J., & Zumbo, B. D. (2016). Validity as a Pragmatist Project: A Global Concern with Local Application. In V. Aryadoust & J. Fox (Eds.), *Trends in Language Assessment Research and Practice* (pp. 555–573). Newcastle: Cambridge Scholars Publishers.

Tadmor, C. T., Tetlock, P. E., & Peng, K. (2009). Acculturation Strategies and Integrative Complexity: The Cognitive Implications of Biculturalism. *Journal of Cross-Cultural Psychology*, *40*(1), 105–139. https://doi.org/10.1177/0022022108326279

Tajfel, H. (1981). *Human Groups and Social Categories: Studies in Social Psychology*. New York: Cambridge University Press.

Taylor, C. (1993). *Reconciling the Solitudes: Essays on Canadian Federalism and Nationalism by Charles Taylor*. McGill-Queen's University Press.

Taylor, C., Appiah, K. A., Habermas, J., Rockefeller, S. C., Walzer, M., & Wolf, S. (1994). *Multiculturalism: Examining the Politics of Recognition* (Expanded Paperback edition; A. Gutmann, Ed.). Princeton, N.J: Princeton University Press.

Todd, S., & Young, A. (1994). *Lives in Hazard* [Documentary]. Retrieved from http://www.imdb.com/title/tt0323082/

Tovar, E., & Simon, M. A. (2010). Factorial Structure and Invariance Analysis of the Sense of Belonging Scales. *Measurement and Evaluation in Counseling and Development*, *43*(3), 199–217. https://doi.org/10.1177/0748175610384811

United Nations. (2018). *Sustainable Cities, Human Mobility and International Migration: A Concise Report*. https://doi.org/10.18356/a11581d8-en

United Nations, Population Division, Department of Economic and Social Affairs (UN DESA). (2017). *Trends in International Migrant Stock: The 2017 revision (United Nations database, POP/DB/MIG/Stock/Rev.2017)*. Retrieved from

http://www.un.org/en/development/desa/population/migration/data/estimates2/data/UN_MigrantStockTotal_2017.xlsx

Van De Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthen, B. (2013). Facing off with Scylla and Charybdis: A comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00770

van Tubergen, F. (2006). *Immigrant Integration: A Cross-National Study*. New York: LFB Scholarly Publishing LLC.

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes* (New Ed edition; M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge: Harvard University Press.

Walton, G. M., & Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology*, *92*(1), 82–96. https://doi.org/10.1037/0022-3514.92.1.82

Walton, G. M., & Cohen, G. L. (2011). A Brief Social-Belonging Intervention Improves Academic and Health Outcomes of Minority Students. *Science*, *331*(6023), 1447–1451. https://doi.org/10.1126/science.1198364

Walzer, M. (1984). *Spheres Of Justice: A Defense Of Pluralism And Equality* (Revised ed. edition). New York: Basic Books.

Wehlage, G. G., Rutter, R. A., & Smith, G. A. (1989). *Reducing the Risks: Schools As Communities of Support*. London ; New York: Falmer Pr.

Weiner, B. (1990). History of Motivational Research in Education. *Journal of Educational Psychology*, *82*(4), 616–622.

Wilder, D. A., & Thompson, J. E. (1980). Intergroup contact with independent manipulations on in-group and out-group interaction. *Journal of Personality and Social Psychology*, *38*(4), 589–603.

Williams, E. J. (1995). Maquiladora Industry and Environmental Degradation in the United States-Mexico Borderlands, The. *St. Mary's Law Journal*, *27*, 765.

Williams, K. D. (2009). Chapter 6 Ostracism. *Advances in Experimental Social Psychology*, *41*, 275–314. https://doi.org/10.1016/S0065-2601(08)00406-1

Wright, M. (2011). Policy Regimes and Normative Conceptions of Nationalism in Mass Public Opinion. *Comparative Political Studies*, *44*(5), 598–624. https://doi.org/10.1177/0010414010396461

Wright, M., & Bloemraad, I. (2012). Is There a Trade-off between Multiculturalism and Socio-Political Integration? Policy Regimes and Immigrant Incorporation in Comparative Perspective. *Perspectives on Politics*, *10*(1), 77–95.

Wu, A. D., Liu, Y., Stone, J. E., Zou, D., & Zumbo, B. D. (2017). Is Difference in Measurement Outcome between Groups Differential Responding, Bias or Disparity? A Methodology for Detecting Bias and Impact from an Attributional Stance. *Frontiers in Education*, *2*. https://doi.org/10.3389/feduc.2017.00039

Wu, A., Li, Z., & Zumbo, B. D. (2007). Decoding the Meaning of Factorial Invariance and Updating the Practice of Multi-group Confirmatory Factor Analysis: A Demonstration With TIMSS Data. *Practical Assessment Research and Evaluation*, *12*, 1–26.

Yildirim, H. H., & Berberoĝlu, G. (2009). Judgmental and Statistical DIF Analyses of the PISA-2003 Mathematics Literacy Items. *International Journal of Testing*, *9*(2), 108–121. https://doi.org/10.1080/15305050902880736

Zimmerman, D., & Zumbo, B. D. (2015). Resolving the Issue of How Reliability is Related to Statistical Power: Adhering to Mathematical Definitions. *Journal of Modern Applied Statistical Methods*, *14*(2). Retrieved from http://digitalcommons.wayne.edu/jmasm/vol14/iss2/5

Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores.* Ottawa, ON: Department of National Defense.

Zumbo, B. D. (2003). Does item-level DIF manifest itself in scale-level analyses? Implications for translating language tests. *Language Testing*, *20*(2), 136–147. https://doi.org/10.1191/0265532203lt248oa

Zumbo, B. D. (2009). Validity as contextualized and pragmatic explanation, and its implications for validation practice. In R. W. Lissitz (Ed.), *The concept of validity* (pp. 65–82). Charlotte, NC: Information Age Publishing Inc.

Zumbo, B. D. (2015, November). *Consequences, Side Effects and the Ecology of Testing: Keys to Considering Assessment 'In Vivo'.* Presented at the Plenary address, the annual meeting of the Association for Educational Assessment – Europe (AEA-Europe), Glasgow, Scotland. Retrieved from http://brunozumbo.com/aea-europe2015/

Zumbo, B. D., & Forer, B. (2011). Testing and measurement from a multilevel view: Psychometrics and validation. In J. A. Bovaird, K. F. Geisinger, C. W. Buckendahl, J. A. Bovaird (Ed), K. F. Geisinger (Ed), & C. W. Buckendahl (Ed) (Eds.), *High-stakes testing in education: Science and practice in K–12 settings.* (pp. 177–190). https://doi.org/10.1037/12330-011

Zumbo, B. D., & Hubley, A. M. (Eds.). (2017). *Understanding and Investigating Response Processes in Validation Research* (1st ed. 2017 edition). New York, NY: Springer.

Zumbo, B. D., Liu, Y., Wu, A. D., Forer, B., & Shear, B. R. (2017). National and International Educational Achievement Testing: A Case of Multi-level Validation Framed by the Ecological Model of Item Responding. In *Social Indicators Research Series*. *Understanding and Investigating Response Processes in Validation Research* (pp. 341–361). https://doi.org/10.1007/978-3-319-56129-5_18

Zumbo, B. D., Liu, Y., Wu, A. D., Shear, B. R., Astivia, O. L. O., & Ark, T. K. (2015). A Methodology for Zumbo's Third Generation DIF Analyses and the Ecology of Item Responding. *Language Assessment Quarterly*, *12*(1), 136–151. https://doi.org/10.1080/15434303.2014.972559

Appendices

Chapter 3 Appendix

Appendix A

Table 3.A Sample Size by Country and Migration Background

	nb with 2 nb parents	nb with 1 fb parent	nb with 2 fb parents	fb	Total
Australia	8,382	2,138	1,332	1,655	13,507
Canada	9,819	1,304	1,577	1,931	14,631
Ireland	3,584	735	176	892	5,387
Lebanon	1,374	147	29	179	1,729
Macao	211	158	271	124	764
Malta	2,767	453	51	202	3,473
New Zealand	2,345	581	450	848	4,224
Qatar	688	176	648	3,050	4,562
Singapore	3,456	1,210	375	958	5,999
Sweden	31	6	10	21	68
Trinidad and Tobago	3,679	268	74	196	4,217
United Arab Emirates	981	214	1,671	3,528	6,394
United Kingdom	9,920	1,086	620	1,319	12,945
United States	3,640	457	886	425	5,408
Total	50,877	8,933	8,170	15,328	83,308

Appendix B

Table 3.B Internal Structure

		CFA			EFA			
Country	Cronbach's Alpha	RMSEA	CFI	TLI	1st Eigenvalue	2nd Eigenvalue	Ratio	Dimensionality
Australia	0.85	0.18	0.95	0.95	3.93	0.8	4.91	essential
Canada	0.87	0.16	0.95	0.92	3.96	0.8	4.95	essential
Ireland	0.86	0.16	0.97	0.95	4.05	0.63	6.43	essential
Lebanon	0.65	0.15	0.82	0.7	2.48	1.33	1.86	2-dimensional
Macao	0.67	0.24	0.76	0.59	2.57	1.37	1.88	2-dimensional
Malta	0.77	0.13	0.95	0.91	3.23	0.87	3.71	2-dimensional
New Zealand	0.83	0.19	0.95	0.91	3.68	0.84	4.38	essential
Qatar	0.80	0.18	0.92	0.87	3.39	0.94	3.61	2-dimensional
Singapore	0.85	0.19	0.95	0.91	3.75	0.81	4.63	essential
Sweden	0.87	0.32	0.95	0.91	4.24	1.03	4.12	essential
Trinidad and Tobago	0.75	0.13	0.93	0.89	3.00	0.97	3.09	2-dimensional
United Arab Emirates	0.78	0.15	0.9	0.83	3.22	1.07	3.01	2-dimensional
United Kingdom	0.84	0.13	0.95	0.92	3.79	0.79	4.80	essential
United States	0.86	0.19	0.95	0.92	3.94	0.77	5.12	essential
Overall	0.83	0.07	0.95	0.92				strict

Appendix C
Mplus Alignment Code

```
  TITLE: 56 group (cntryXmiggroup). Free estimated.
    DATA: FILE = Paper1v2.dat;
     VARIABLE: NAMES = item1 item2 item3 item4 item5 item6 weight country
mig_cnt;
      Categorical= item1-item6;
      USEVARIABLES = item1-item6 mig_cnt;
      MISSING ARE ALL (99);
      WEIGHT= weight;
      CLASSES= c1(56);
      KNOWNCLASS = c1(mig_cnt= 1-56);
      ANALYSIS: TYPE = MIXTURE;
      ESTIMATOR = MLR;
      PROCESSORS = 2;
      ALIGNMENT = FREE;
      ALGORITHM=INTEGRATION;
      STARTS = 2000 50;
      MODEL: %OVERALL%
       f BY item1-item6;
       OUTPUT: TECH1 TECH8 ALIGN;
       PLOT: TYPE = PLOT2;
```

Appendix D

Table 3.C PISA Social Belonging: Approximate (Non-) Invariance for Intercepts over 14 countries and 4 migration backgrounds

Country	Mig Group	1$1	1$2	1$3	2$1	2$2	2$3	3$1	3$2	3$3	4$1	4$2	4$3	5$1	5$2	5$3	6$1	6$2	6$3	# Noninvariant
Qatar	heritage	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Qatar	nb 1fb	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0
Qatar	**nb 2fb**	(3)	3	(3)	3	3	(3)	3	3	3	3	3	(3)	(3)	3	3	(3)	3	(3)	**7**
Qatar	**fb**	(4)	4	(4)	4	4	(4)	4	4	4	(4)	4	(4)	(4)	(4)	4	4	4	4	**7**
Australia	heritage	5	5	5	5	5	5	5	(5)	(5)	5	5	5	5	5	(5)	5	5	5	3
Australia	nb 1fb	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	0
Australia	nb 2fb	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	0
Australia	fb	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	0
Singapore	heritage	6	6	6	6	6	6	6	6	6	6	6	6	(6)	(6)	(6)	6	6	6	3
Singapore	nb 1fb	7	7	7	7	7	7	7	7	7	7	7	7	7	(7)	7	7	7	7	1
Singapore	nb 2fb	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	0
Singapore	fb	9	9	9	9	9	9	9	9	9	9	9	9	9	(9)	9	9	9	9	1
Sweden	heritage	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	0
Sweden	nb 1fb	12	12	12	12	12	12	(12)	12	12	12	12	12	12	12	12	12	12	12	1
Sweden	nb 2fb	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	0
Sweden	fb	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	0
Trinidad and Tobago	**heritage**	(16)	16	(16)	16	(16)	(16)	(16)	16	(16)	(16)	16	(16)	(16)	(16)	(16)	(16)	16	(16)	**13**
Trinidad and Tobago	nb 1fb	17	17	17	17	17	(17)	(17)	17	17	17	17	17	17	17	(17)	17	17	17	3
Trinidad and Tobago	nb 2fb	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	0
Trinidad and Tobago	**fb**	19	19	(19)	19	19	19	19	19	(19)	19	19	(19)	19	19	19	(19)	19	(19)	**5**
United Arab Emirates	**heritage**	21	21	(21)	21	21	(21)	(21)	21	(21)	21	21	21	(21)	21	(21)	21	21	(21)	**7**
United Arab Emirates	nb 1fb	22	22	22	22	22	22	(22)	22	22	22	22	22	22	22	22	22	22	22	1
United Arab Emirates	**nb 2fb**	(23)	23	(23)	23	23	(23)	(23)	23	(23)	(23)	23	(23)	(23)	23	(23)	(23)	23	(23)	**11**
United Arab Emirates	**fb**	(24)	24	(24)	(24)	24	(24)	24	24	24	24	24	(24)	(24)	24	24	24	24	(24)	**7**
United Kingdom	**heritage**	25	25	25	25	25	(25)	(25)	(25)	(25)	25	25	25	25	25	(25)	25	(25)	25	**6**
United Kingdom	nb 1fb	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	(26)	1
United Kingdom	nb 2fb	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	0
United Kingdom	fb	28	28	28	28	28	28	28	28	28	28	28	28	(28)	28	28	28	28	28	1
United States	heritage	29	29	29	29	29	29	29	29	29	29	29	29	29	(29)	29	29	29	29	1
United States	nb 1fb	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	0
United States	nb 2fb	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	0
United States	fb	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	0
Canada	heritage	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	0
Canada	nb 1fb	34	34	34	34	34	34	34	34	34	34	34	34	34	34	34	34	34	34	0
Canada	nb 2fb	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	0
Canada	fb	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	0
Ireland	**heritage**	37	(37)	37	37	37	(37)	(37)	(37)	(37)	37	37	37	37	37	(37)	37	37	37	**6**
Ireland	nb 1fb	38	38	38	38	38	(38)	38	38	38	38	38	38	38	38	(38)	38	38	38	2
Ireland	nb 2fb	39	39	39	39	39	39	39	39	39	39	39	39	39	39	39	39	39	39	0
Ireland	fb	40	40	40	40	40	40	40	40	(40)	40	40	40	40	40	40	40	40	40	1
Lebanon	**heritage**	(41)	41	(41)	41	(41)	(41)	41	(41)	(41)	41	41	(41)	(41)	41	(41)	(41)	41	(41)	**11**
Lebanon	nb 1fb	42	42	(42)	42	42	42	42	42	42	42	42	(42)	42	42	42	42	42	(42)	3

Lebanon	nb 2fb	43	43	43	43	43	43	43	43	43	43	43	43	43	43	43	43	43	43	0
Lebanon	fb	44	44	44	44	44	44	44	44	44	44	44	44	44	44	44	44	44	(44)	1
Macao	heritage	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	0
Macao	nb 1fb	46	46	46	46	46	46	46	46	46	46	46	46	46	46	46	46	46	46	0
Macao	nb 2fb	47	47	47	47	47	47	47	47	47	47	47	47	(47)	(47)	47	47	47	47	2
Macao	fb	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	0
Malta	**heritage**	49	49	(49)	49	49	49	(49)	(49)	49	49	49	(49)	49	49	49	49	49	(49)	**5**
Malta	nb 1fb	50	50	50	50	50	50	(50)	(50)	50	50	50	50	50	50	50	50	50	(50)	3
Malta	nb 2fb	51	51	51	51	51	51	51	51	51	51	51	51	51	51	51	51	51	51	0
Malta	fb	52	52	52	52	52	52	(52)	(52)	52	52	52	(52)	52	52	52	52	52	52	3
New Zealand	heritage	53	53	53	53	53	53	(53)	53	53	53	53	53	53	53	(53)	53	53	53	2
New Zealand	nb 1fb	54	54	54	54	54	54	54	54	54	54	54	54	54	54	54	54	54	54	0
New Zealand	nb 2fb	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55	0
New Zealand	fb	56	56	56	56	56	56	56	56	56	56	56	56	56	56	56	56	56	56	0

Note: Columns are - Item$Threshold; **Bold** country groups represent those groups that exceed 25% of (non)invariance and are not comparable.

Appendix E

Table 3.D Approximate (Non-) Invariance for Loadings over 14 countries and 4 migration backgrounds

Country	Mig Group	ITEM1	ITEM2	ITEM3	ITEM4	ITEM5	ITEM6
Qatar	heritage	1	1	1	1	1	1
Qatar	nb 1fb	2	2	2	2	(2)	2
Qatar	**nb 2fb**	(3)	(3)	(3)	(3)	(3)	(3)
Qatar	fb	4	4	4	4	4	4
Australia	heritage	5	5	5	5	5	5
Australia	**nb 1fb**	(10)	(10)	(10)	(10)	(10)	(10)
Australia	nb 2fb	15	15	15	15	15	15
Australia	fb	20	20	20	20	20	20
Singapore	heritage	6	6	6	6	6	6
Singapore	nb 1fb	7	7	7	7	7	7
Singapore	nb 2fb	8	8	8	8	8	8
Singapore	fb	9	9	9	9	9	9
Sweden	heritage	11	11	11	11	11	11
Sweden	nb 1fb	12	12	12	12	12	12
Sweden	nb 2fb	13	13	13	13	13	13
Sweden	**fb**	(14)	14	(14)	(14)	14	14
Trinidad and Tobago	**heritage**	(16)	(16)	(16)	(16)	(16)	(16)
Trinidad and Tobago	**nb 1fb**	(17)	17	17	(17)	17	(17)
Trinidad and Tobago	nb 2fb	18	18	18	18	18	18
Trinidad and Tobago	fb	19	19	19	19	19	19
United Arab Emirates	heritage	21	21	21	21	21	21
United Arab Emirates	nb 1fb	22	22	(22)	22	22	22
United Arab Emirates	nb 2fb	23	23	23	23	23	23
United Arab Emirates	fb	24	24	24	24	24	24
United Kingdom	heritage	25	25	25	25	25	25
United Kingdom	nb 1fb	26	26	26	26	26	26
United Kingdom	nb 2fb	27	27	27	27	27	27
United Kingdom	fb	28	28	28	28	28	28
United States	**heritage**	(29)	(29)	(29)	(29)	(29)	(29)
United States	nb 1fb	30	30	30	30	30	30
United States	nb 2fb	31	31	31	31	31	31
United States	fb	32	32	32	32	32	32
Canada	heritage	33	33	33	33	33	33
Canada	nb 1fb	34	34	34	34	34	34
Canada	nb 2fb	35	35	35	35	35	35
Canada	fb	36	36	36	36	36	36
Ireland	**heritage**	(37)	(37)	(37)	(37)	(37)	(37)
Ireland	nb 1fb	38	38	38	38	38	38
Ireland	nb 2fb	39	39	39	39	39	39
Ireland	fb	40	40	40	40	40	40
Lebanon	heritage	41	41	41	41	41	41
Lebanon	nb 1fb	42	42	42	42	42	42
Lebanon	nb 2fb	43	43	43	43	43	43
Lebanon	fb	44	44	44	44	44	44
Macao	heritage	45	45	45	45	45	45
Macao	nb 1fb	46	46	46	46	46	46
Macao	**nb 2fb**	(47)	(47)	47	47	(47)	(47)

Macao	fb	48	48	48	48	48	(48)
Malta	**heritage**	(49)	(49)	(49)	(49)	(49)	(49)
Malta	**nb 1fb**	(50)	(50)	(50)	(50)	(50)	(50)
Malta	nb 2fb	51	51	51	51	51	51
Malta	fb	52	52	52	52	52	52
New Zealand	heritage	53	53	53	53	53	53
New Zealand	nb 1fb	54	54	54	54	54	54
New Zealand	nb 2fb	55	55	55	55	55	55
New Zealand	fb	56	56	56	56	56	56

Note: **Bold** country groups represent those groups that exceed 25% of (non)invariance and are not comparable.

Chapter 4 Appendix

Appendix A

Table 4.A1 Cross Tab of Language and Country (Unweighted Counts)

	Basque	Bokmål	Bulgarian	Catalan	Croatian	Czech	Danish	Dutch	English	Estonian	Finnish	French	Galician	German	Greek	Hungarian
Australia									13,776							
Austria														6,931		
Belgium								5,515				3,479		378		
Bulgaria			5708													
Canada									14,725			4,273				
Croatia					5624											
Czech Republic						6792										
Denmark							6,981									
Estonia										4,268						
Finland											5,458					
France												5,948				
Germany														5,696		
Greece															5,425	
Hungary																5,567
Iceland																
Ireland									5,397							
Italy														1,545		
Japan																
Korea																
Latvia																
Lithuania																
Malta									3,492							
Netherlands								5,213								
New Zealand									4,264							
Norway		4832														
Poland																
Portugal																
Romania																414
Slovak Republic																393
Slovenia																
Spain	1300			4,184									1,556			
Sweden									68							
Switzerland												1,289		3,443		
Turkey																
United Kingdom									13,084							
United States									5,474							
Total	1,300	4,832	5,708	4,184	5,624	6,792	6,981	10,728	60,280	4,268	5,458	14,989	1,556	17,993	5,425	6,374

Table 4.A2 Cross Tab of Language and Country (Unweighted Counts)

	Icelandic	Italian	Japanese	Korean	Latvian	Lithuanian	Polish	Portuguese	Romanian	Russian	Slovak	Slovenian	Spanish	Swedish	Turkish	Total
Australia																13,776
Austria																6,931
Belgium																9,372
Bulgaria																5,708
Canada																18,998
Croatia																5,624
Czech Republic																6,792
Denmark																6,981
Estonia										1,194						5,462
Finland														340		5,798
France																5,948
Germany																5,696
Greece																5,425
Hungary																5,567
Iceland	3,257															3,257
Ireland																5,397
Italy		9,606										84				11,235
Japan			6,598													6,598
Korea				5,537												5,537
Latvia					3,543					1,256						4,799
Lithuania						4,947	591			725						6,263
Malta																3,492
Netherlands																5,213
New Zealand																4,264
Norway																4,832
Poland							4,429									4,429
Portugal								7,185								7,185
Romania									4,404							4,818
Slovak Republic											5,775					6,168
Slovenia												6,300				6,300
Spain													30,820			37,860
Sweden														5,218		5,286
Switzerland		1,016														5,748
Turkey															5,744	5,744
United Kingdom																13,084
United States																5,474
Total	3,257	10,622	6,598	5,537	3,543	4,947	5,020	7,185	4,404	3,175	5,775	6,384	30,820	5,558	5,744	271,061

Appendix B

Table 4.B MIPEX Policy Domains and sub-scales

	AT	AU	BE	BG	CA	HR	CY	CZ	DK	EE	FI	FR	DE	GR	HU	IS	IE	IT	JP	KR	LV	LT	LU	MT	NL	NO	NZ	PL	PT	RO	SK	SI	ES	SE	CH	TU	UK	US
Labour market mobility	64	58	64	50	81	54	34	52	79	73	80	54	86	55	40	51	38	66	65	71	46	40	42	45	73	90	67	38	91	57	21	38	72	98	59	15	56	67
Access	60	70	60	80	90	70	0	80	80	40	100	20	70	70	60	50	40	90	80	70	30	60	20	60	90	80	70	70	100	50	0	30	100	90	60	0	80	100
General support	42	50	92	50	83	50	90	58	58	92	67	58	83	75	17	50	8	67	67	58	83	42	67	42	83	92	58	33	83	50	33	50	83	100	58	42	75	75
Targeted support	80	60	30	20	50	20	10	20	90	60	80	50	90	0	20	30	30	20	40	70	10	10	30	30	20	90	90	0	80	40	0	20	30	100	30	17	20	30
Workers' rights	75	50	75	50	100	75	38	50	88	100	75	88	100	75	63	75	75	88	75	88	63	50	50	50	100	100	50	50	100	88	50	50	75	100	88	0	50	63
Family reunion	50	67	72	64	79	69	39	57	42	67	68	51	57	55	61	59	40	72	61	64	55	59	65	48	56	63	68	65	88	67	56	80	90	78	48	49	33	66
Eligibility	46	54	64	61	64	61	7	50	7	43	57	39	25	36	89	79	36	71	75	71	64	39	64	25	43	57	79	57	93	71	71	79	79	79	54	71	21	68
Conditions	29	70	61	60	70	80	50	50	46	80	80	23	50	50	60	60	60	44	80	75	60	80	70	51	48	46	60	60	70	60	50	90	80	70	31	70	19	70
Security	40	60	80	60	80	50	50	60	40	70	60	60	70	60	60	30	40	80	30	50	20	40	50	40	60	50	50	60	90	60	60	70	100	70	50	20	50	60
Rights	83	83	83	75	100	83	50	67	75	75	75	83	83	75	33	67	25	92	58	58	75	75	75	75	75	100	83	83	100	75	42	83	100	92	58	33	42	67
Education	47	76	61	3	65	15	27	38	49	58	60	36	47	36	15	23	30	34	21	57	17	17	48	19	50	65	66	20	62	20	24	26	37	77	42	5	57	60
Access	58	58	42	0	50	8	33	33	42	50	67	58	50	50	0	17	25	25	25	75	17	8	42	17	50	42	58	17	67	8	17	33	42	58	42	8	42	83
Targeting needs	43	87	65	0	80	13	53	60	90	90	90	30	47	23	30	37	47	60	40	73	17	30	57	20	60	90	87	23	70	63	20	17	37	90	47	3	67	90
New opportunities	35	80	53	0	60	0	0	20	45	40	45	25	50	30	30	0	0	10	10	40	15	0	25	0	10	50	38	20	40	0	0	15	30	80	40	0	30	15
Intercultural education	50	80	85	10	70	40	20	40	20	50	40	30	40	40	0	40	50	40	10	40	20	30	70	40	80	80	80	20	70	10	60	40	40	80	40	10	90	50
Political participation	38	64	57	13	48	13	25	21	64	21	79	53	63	30	23	67	73	58	31	54	13	16	81	25	52	82	74	6	74	0	16	23	54	71	58	11	51	36
Electoral rights	0	25	13	0	0	0	0	0	75	17	67	0	0	0	17	67	67	0	0	25	0	33	67	0	50	75	75	0	50	0	38	17	25	75	38	0	50	0
Political liberties	100	100	100	50	100	50	100	50	100	25	100	100	100	100	75	75	100	100	100	50	25	25	100	100	100	100	100	25	100	0	25	50	100	100	100	25	100	100
Consultative bodies	20	40	25	0	0	0	0	13	50	7	68	33	63	20	0	60	37	53	25	40	0	7	70	0	20	63	20	0	48	0	0	0	53	10	45	8	25	15
Implementation policies	30	90	90	0	90	0	0	20	30	38	80	80	90	0	0	67	88	80	0	100	25	0	88	0	38	90	100	0	100	0	0	25	40	100	50	10	30	30
Permanent residence	57	54	86	67	62	65	37	51	74	71	70	48	60	54	68	62	49	65	59	54	53	59	64	50	55	70	64	66	68	57	54	61	74	79	51	27	51	54
Eligibility	50	75	63	63	63	50	25	50	88	50	50	13	50	63	50	63	13	38	38	38	63	50	50	25	25	50	75	50	50	50	50	38	63	75	38	25	38	50
Conditions	29	50	100	50	29	40	7	29	60	71	67	31	33	2	67	43	67	55	50	67	19	40	67	12	40	60	62	50	55	33	24	67	67	67	17	50	10	67
Security	50	56	81	56	56	69	50	44	50	63	63	81	56	50	56	44	19	69	50	44	31	44	38	63	56	69	52	63	69	44	44	75	69	75	50	31	56	31
Rights	100	33	100	100	100	100	67	83	100	100	100	67	100	100	100	100	100	100	100	67	100	100	100	100	100	100	67	100	100	100	100	67	100	100	100	0	100	67
Access to nationality	26	69	69	21	67	31	37	49	58	18	63	61	72	34	31	53	59	50	37	36	17	35	68	34	66	52	71	56	86	34	25	41	48	73	31	34	60	61
Eligibility	21	83	63	0	79	33	42	33	63	8	79	79	92	29	4	58	96	42	46	42	0	25	58	38	71	38	67	20	92	29	8	21	50	50	38	50	79	75
Conditions	17	50	28	50	67	42	45	28	42	42	58	25	52	37	43	37	33	23	48	48	33	58	75	45	42	88	68	63	78	15	15	53	32	83	2	23	27	37
Security	30	43	83	23	23	50	10	60	27	20	40	40	57	20	27	40	7	60	30	43	10	30	63	3	63	43	50	90	73	40	27	53	60	60	33	13	33	33
Dual nationality	38	100	100	13	100	0	50	75	100	0	75	100	88	50	50	75	100	75	25	13	25	25	75	50	88	38	100	50	100	50	50	38	50	100	50	50	100	100
Anti-discrimination	57	74	78	89	92	61	50	48	50	32	77	77	58	60	83	5	66	61	22	52	34	43	49	51	73	59	79	52	88	78	72	67	49	85	31	26	85	90
Definitions	67	67	75	92	100	67	42	58	50	42	75	67	67	58	67	0	58	58	17	33	33	58	50	50	67	58	75	50	75	67	67	67	50	75	58	42	100	100
Fields of application	60	70	100	100	100	50	50	50	50	10	100	100	80	50	100	0	100	100	30	100	30	20	50	50	80	50	80	50	100	100	100	100	50	100	0	20	100	100
Enforcement mechanisms	56	88	75	94	75	81	63	69	63	44	56	81	69	75	94	19	50	75	31	31	38	44	75	75	94	56	81	88	94	94	81	63	69	81	38	25	69	94
Equality policies	44	72	61	72	94	44	44	17	39	33	78	61	17	56	72	0	56	11	11	44	33	50	22	28	50	72	78	22	83	50	39	39	28	83	28	17	72	67
Health	63	67	53	28	49	20	31	44	53	27	53	50	43	27	40	40	58	65	51	36	17	26	43	45	55	67	75	26	43	45	31	18	53	62	70	32	64	69
Entitlement to health services	64	56	69	50	42	53	33	58	58	39	56	83	50	61	53	61	53	72	75	22	31	31	69	28	78	69	69	64	33	67	50	36	50	78	78	72	39	44
Policies to facilitate access	62	53	72	28	53	20	48	53	48	45	73	82	30	28	35	80	55	78	90	68	28	40	52	58	55	60	70	38	70	55	43	25	67	62	85	32	42	73
Responsive health services	71	75	42	0	50	0	17	29	46	0	50	25	58	0	42	21	58	50	21	17	0	8	42	46	46	58	83	0	29	25	0	4	38	58	63	4	92	79
Measures to achieve change	54	83	29	33	50	8	25	33	58	25	33	8	33	17	29	0	67	58	17	38	8	25	8	50	42	79	79	0	38	33	29	8	58	50	54	21	83	79

Appendix C

Table 4.C Count of Language-Based Non-Invariant Parameters

Language	Language Group	Threshold	Intercept
Basque	Vasconic	3	1
Icelandic	Germanic	3	2
Russian	Balto-Slavic	4	2
Slovak	Balto-Slavic	4	3
Slovenian	Balto-Slavic	4	2
Croatian	Balto-Slavic	5	2
Galician	Romance	5	2
Hungarian	Finnic	6	3
Czech	Balto-Slavic	7	5
Finnish	Finnic	7	2
Polish	Balto-Slavic	7	2
Bokmål	Germanic	8	3
Catalan	Romance	8	6
Danish	Germanic	8	3
Latvian	Balto-Slavic	8	4
Portuguese	Romance	8	1
Swedish	Germanic	8	5
Greek	Hellenic	9	4
Korean	Koreanic	10	6
Dutch	Germanic	11	2
Italian	Romance	11	3
Japanese	Japonic	11	3
Bulgarian	Balto-Slavic	12	2
Estonian	Finnic	12	6
Lithuanian	Balto-Slavic	13	3
Romanian	Romance	13	3
French	Romance	14	4
German	Germanic	14	4
Spanish	Romance	14	5
Turkish	Turkic	14	3
English	Germanic	16	6

Note: Languages with more than 6 non-invariant parameters are considered not comparable

Appendix D

Figure 4.1 Q-Q Plot of Country-Level Residuals