

Boston College
Lynch School of Education and Human Development

Department of
Measurement, Evaluation, Statistics, and Assessment

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION
IN K-12 SCHOOLS IN THE UNITED STATES: A SCALE DEVELOPMENT USING
RASCH/GUTTMAN SCENARIO METHODOLOGY

Dissertation
by
MARÍA EUGENIA BÁEZ CRUZ

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

May 2021

ABSTRACT

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION IN K-12 SCHOOLS IN THE UNITED STATES: A SCALE DEVELOPMENT USING RASCH/GUTTMAN SCENARIO METHODOLOGY

María Eugenia Báez Cruz, Author

Larry H. Ludlow, Chair

In 2019, as in previous years, White students outperformed African American, Hispanic, and American Indian/Alaska Natives in a variety of K-12 outcomes (Darling-Hammond, 2007; de Brey et al., 2019; Jacob & Ludwig, 2008; National Education Association, 2015). The urgency of the opportunity/achievement gap is clear, as the current cohort of students under 5 years of age marks a turning point in student population demographics as the first in which 50 percent are part of a minority race or ethnic group (U.S. Census Bureau, 2015).

Sociocultural integration (SCI) is included in the frameworks of successful bilingual programs (Howard et al., 2007; Scanlan & López, 2014). SCI considers the dynamics of relationships with oneself and others as being built in the context of one's racial/ethnic, cultural, and linguistic background (Brisk, 2006; Feinauer & Howard, 2014). Acceptance and appreciation of cultural difference is critical for teachers (Bennett, 2003) and a number of pedagogical frameworks center teachers' role of cultural brokerage as a pathway to fostering positive student outcomes (Grant & Sleeter, 2006; Suárez-Orozco & Suárez-Orozco, 2001; Villegas & Lucas, 2002).

In this dissertation, I defined sociocultural integration in a teacher-centered way, and explicitly incorporate teachers' racial/ethnic identity development in the evolution of

their actions to support SCI. Second, I operationalized this definition and built a scale for measuring SCI using innovative “lived experiences” scenario items according to the Rasch/Guttman Scenario scale development methodology (Ludlow et al., 2020). The SCI Scale for Teachers showed desirable psychometric properties and is well suited to increase use due to ease of interpretability.

ACKNOWLEDGEMENTS

This dissertation has been a true labor of love. Completed during the COVID-19 pandemic, it was a test of perseverance, grit, and love. As a new mom, a full-time employee, and an international student with her whole family abroad, the continuous support of those I love and admire were what powered me through writing and sacrificing time with my firstborn to complete my doctorate. I dedicate my dissertation to my son, Alfonso Javier, and my husband, Diego. To Alfonso, for inspiring me to be a better human being and teaching me that the biggest joys of life happen far from a computer screen. To Diego, my love, for supporting me beyond measure, like only the best life partners do. Your discipline is an example I try to live by. Through this challenging year, we have emerged as a strong team that can weather any test.

I want to thank my committee chair, Dr. Larry Ludlow, for supporting me and being adaptable. For giving me deadlines, holding me to my word, and questioning me. Academic rigor is only achieving through such challenges. Dr. Ludlow's kind and joking manner made the process of writing this dissertation seem a less insurmountable task. I am also extremely grateful to Dr. Patrick Proctor, whose guidance was critical to every aspect of this dissertation. Dr. Proctor always made time to discuss any scale review or literature and provided actionable feedback on every occasion. He eased my impostor syndrome throughout. Finally, I want to thank Dr. Martin Scanlan. Dr. Scanlan gave me the confidence to turn a topic I was passionate about into a dissertation research question. He showed me how to collaborate with other academics and encouraged me to reach across fields of study. Dr. Scanlan put his own network at my disposal to explore data collection sites and partners. The group-think and problem solving I was able to engage

in with my committee convinced me that academic work that stands the scrutiny of scholars one respects is worthy and fulfilling. I cannot imagine a life away from research.

Persistence is an exercise one must engage in every day. I would not have been able to do so successfully without my mother, Carmen. Mama, our daily conversations, your visit at the most critical of times, nurtured my soul. My father, Julio, and my brother, Julio Miguel, were fuel for my typing. Always ensuring me they were proud and that they trusted me to complete my dissertation without compromising my commitment to excellence. I want to thank my best friend, Nilka, who let me break down and helped me piece myself back together – more than once – but to a stronger version of myself every time.

Finally, to the tribe of women scholars, friends, and collaborators who have inspired me throughout the PhD: thank you. Thank you, Tolani Britton, for supporting me since we met when I was master's student. A truly excellent human being in all dimensions, you continue to show me that I, and other women of color, can be scholars, mothers, and partners – and do all this gracefully. I am so grateful to Gabrielle Oliveira, who trusted me as a young researcher and empowered me to make decisions for our project. You showed me what an ideal research group looks like, and how we can lift others by engaging in quality work that motivates everyone to do their best. Thank you, Cristina Hunter, for being an example for me, a strong Latina researcher, and for being there for me as a true friend. Chris Chang-Bacon, I thank you for your incredible feedback, for being such a generous research collaborator, and for showing me that we can be kind and excellent. Thank you Kelsey Klein and Caroline Vuilleumier for being the wind beneath my wings. For motivating me, holding me accountable, for listening to

me complain, and giving me important feedback for my defense. To Ana Tavares, thank you for your friendship, encouragement, and advice.

Finally, I want to express my immense gratitude to the TWIN-CS community, where my research question was born and nurtured. To Patty, for her unbelievable support since day one. To Mary Bridget Burns, Stephanie Margetts, and Myra Reynoso, for opening the doors to work with teachers and administrators of the network who I greatly admire and respect. I also want to thank star administrator Michael Guerra, whose feedback was critical for the development of the SCI-T.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION	1
Deepening of the Opportunities to Learn Gap	1
Motivation of the Study	3
Purpose of the Study	8
Significance of the Study	11
Organization of the Dissertation	13
CHAPTER 2 LITERATURE REVIEW	15
Sociocultural Integration: A Construct of Interest for Equitable Opportunities to Learn	16
Self-identification: Racial/Ethnic and Cultural Identity as an Anchor for Teachers' Promotion of Sociocultural Integration.....	22
Racial Identity Models	25
Cultural Identity Models	31
Frameworks Promoting Healthy Identity Development in Students and Teacher Agency to Pursue Equity	36
Measurement of Racial/Ethnic and Cultural Identity Development, Healthy Identity Promotion, and Agency for Equity.....	44
Proposed Model for Measuring Teacher Promotion of Sociocultural Integration	51
CHAPTER 3 RESEARCH METHODOLOGY	57

Using the Rasch/Guttman Scenario Scale Methodology to Improve Score	
Usability in the Measurement of Sociocultural Integration	57
Rasch/Guttman Scenario Scale Development Methodology	58
Research Design.....	72
Population of Interest.....	74
Step One: Defining Sociocultural Integration.....	75
Steps Two and Three: Operationalizing SCI Using Facets, and Determining Levels	78
Steps Four and Five: Formulating Sentence Mapping Structures and Constructing Scenarios	80
Step Six: Selecting Response Options and Survey Instructions	84
Step Seven: Testing Congruence of SCI Theory and Practice	88
CHAPTER 4 RESULTS	97
Pilot Study.....	98
Pilot Study Missing Data	98
Pilot Study Scenario Descriptive Statistics.....	100
Pilot Study Variable Map (n=68).....	102
SCI-T Revisions informed by Pilot Study Conclusions	115
Main Study.....	116
Main Study Missing Data	116
Main Study Sample and Scenario Descriptive Statistics	117

Main Study Variable Map.....	121
Main Study Rating Scale Model Results	123
Main Study Differential Item Functioning Analysis.....	141
SCI-T Score Interpretation.....	147
CHAPTER 5 DISCUSSION.....	152
Summary of Findings.....	152
Study Implications	155
Areas of Future Research.....	158
REFERENCES	160
APPENDIX 1.....	187
APPENDIX 2.....	188
APPENDIX 3.....	189
APPENDIX 4.....	192
APPENDIX 5.....	202
APPENDIX 6.....	203
APPENDIX 7.....	207
APPENDIX. 8.....	211
APPENDIX. 9.....	212
APPENDIX. 10.....	213

TABLES

Table 2.1	Stages of Self-Identification for Teacher Promotion of SCI.....	36
Table 3.1	Response Options Considered and Tested for SCI-T	86
Table 3.2	Survey Instructions Tested for SCI-T	88
Table 3.3	SCI-T Design for Cognitive Lab.....	90
Table 3.4	Facet and Level Combinations Used in Pilot Study Scenarios.....	94
Table 4.1	Pilot Study: Scale Completion Rates	99
Table 4.2	Pilot Study: Responses by Survey Language.....	100
Table 4.3	Pilot Study: Scenario Descriptive Statistics.....	101
Table 4.4	Pilot Study: Difficulty Estimates and Fit Statistics.....	106
Table 4.5	Pilot Study: PCA of Residuals Analysis – Total Variance Explained.....	114
Table 4.6	Main Study: Scale Completion Rates	117
Table 4.7	Main Study: Teacher Characteristics	119
Table 4.8	Main Study: Scenario Descriptive Statistics	120
Table 4.9	Main Study: Difficulty Estimates and Fit Statistics.....	124
Table 4.10	Main Study: Selected Respondent Fit Statistics (n=14).....	125
Table 4.11	Main Study: Model Fit Statistics Before and After Deletion.....	127
Table 4.12	Main Study: Difficulty Estimates and Fit Statistics After Deletion.....	131
Table 4.13	Main Study: PCA of Responses – Variance Explained	135
Table 4.14	Main Study: PCA of Residuals Analysis – Total Variance Explained	137
Table 4.15	Main Study: Residual Analysis - Rotated Component Matrix.....	138
Table 4.16	Main Study: Uniform DIF Analysis, by Ethnoracial Categories	145
Table 4.17	Main Study: Uniform DIF Analysis Summary	146

Table 4.18 Main Study: Score to SCI Level Interpretation.....	149
--	-----

FIGURES

Figure 1.1 Conceptual Model for Definition of Sociocultural Integration	7
Figure 2.1 Proposed Model for Teachers’ Promotion of Sociocultural Integration	22
Figure 2.2 Relationship between Facets in Model for Teacher Promotion of SCI.....	53
Figure 3.1 Diagram of the RGS Scenario Development Process	73
Figure 4.1 Pilot Study: SCI-Teacher Scale Variable Map (n=68)	103
Figure 4.2 Pilot Study: Scenario Response Category Characteristic Curves.....	109
Figure 4.3 Pilot Study: Residual Matrix Principal Component Analysis Scree Plot.....	111
Figure 4.4 Main Study: Variable Map (n=163)	122
Figure 4.5 Main Study: Variable Map After Deletion (n=156)	129
Figure 4.6 Main Study: Cumulative Probability Map by Andrich Thresholds.....	133
Figure 4.7 Main Study: Comparison of Non-Hispanic Whites to Hispanic, Black, Asian and Two-or-more Category Respondents	140
Figure 4.8 Main Study: Estimated Scenario Difficulties, by Ethnoracial Categories....	142
Figure 4.9 Main Study: SCI-T Interpretation by Levels of Enactment	148

CHAPTER 1 INTRODUCTION

Deepening of the Opportunities to Learn Gap

In 2019, as in previous years, White students outperformed African American, Hispanic, and American Indian/Alaska Natives in a variety of K-12 outcomes (Darling-Hammond, 2007; de Brey et al., 2019; Jacob & Ludwig, 2008; National Education Association, 2015). The inability of the United States to raise all students to satisfactory academic performance is worrisome. The urgency of the opportunity/achievement gap is clear. The current cohort of students under 5 years of age marks a turning point in student population demographics as the first in which 50 percent are part of a minority race or ethnic group (Colby & Ortman, 2015), including 5 million English language learners who accounted for 10.1 percent of public school students nationwide in 2017 (de Brey et al., 2019).

At the same time, the United States is currently engaged in a tumultuous racial reckoning in light of despicable displays of racism towards Black, brown, and Indigenous people; police brutality towards Black men and women that has gone unpunished; and astounding rates of infection and death from COVID-19 for Black, brown, and Indigenous individuals, coupled with the vulnerability of their families to pandemic-associated economic hardships (ACLU, 2020; Chang et al., 2020; Gomez et al., 2020). Pre-existing educational inequalities have also deepened during the pandemic, and the lack of accurate statistics should not convince us otherwise. For example, a sizeable population of the country's most vulnerable students were not assessed in fall 2020 (Kuhfeld et al., 2020), leading us to likely underestimate test score and other impacts of COVID-19. We do know, though, that attrition has been disproportionately concentrated

among ethnic/racial minority students and students whose schools serve more students with socioeconomic disadvantage (Kuhfeld et al., 2020). These students are likely missing virtual or in-person classes because they lacked reliable technology to participate in remote assessments or experienced economic, health, or other hardship.

The growing percentage of non-White students, coupled with the deepening of educational inequities because of the COVID-19 pandemic, reinforces the need to focus on equitable opportunities to learn. As Kendi (2019) states: “there is no neutrality in the racism struggle ... one either allows racial inequities to persevere, as a racist, or confronts racial inequities, as an antiracist” (p. 9). Educational equity is not merely a value-laden goal. There would be significant economic gains to eliminating educational opportunity/achievement gaps (Auguste et al., 2009). To reduce gaps in educational outcomes, many scholars propose the overturn of pre-existing structural inequalities such as unequal access to educational opportunities and resources to learn (e.g., Baker et al., 2016; Ladson-Billings, 2006; Sleeter, 2013; Welner & Farley, 2010).

Multicultural education (Sleeter & Grant, 1987), culturally responsive teaching (Gay, 2000; Ladson-Billings, 1995; Villegas & Lucas, 2002), and social justice teaching (Cochran-Smith, 1999, 2010; Cochran-Smith et al., 2016) are educational frameworks that draw from and celebrate racial, ethnic, cultural, and linguistic diversity – in addition to and beyond academic achievement – to both situate and equip students as critical members of society. While linked to these educational frameworks, sociocultural integration (SCI) is explicitly addressed as a goal of bilingual education (Brisk, 2006; Feinauer & Howard, 2014). In this literature, the concept of sociocultural integration considers relationships with oneself and others as dynamics built through the lens of a

person's racial, ethnic, cultural, and linguistic background. While pedagogical frameworks supporting equity-centered and socially-just teaching are substantial, there is little understanding about how teachers interpret and enact this kind of teaching practice in classrooms (Chang, 2017). In this dissertation, I focus on understanding how teachers enact SCI in their daily practice, as informed by racial/ethnic and cultural identity development models and selected pedagogical frameworks (multicultural education, culturally responsive teaching, and social justice teaching).

Motivation of the Study

Sociocultural integration appears in the educational literature in the realm of bilingual education. Brisk (2006) defines it as the ability to transit between one's heritage community and a society that also includes others' heritage communities. A concept included in the frameworks of some notably successful bilingual programs (Howard et al., 2007; Scanlan & López, 2014), SCI considers the dynamics of relationships with oneself and others as being built in the context of one's racial, ethnic, cultural, and linguistic background (Brisk, 2006; Feinauer & Howard, 2014), as mentioned. Acceptance and appreciation of cultural difference is critical for teachers (Bennett, 2003) and a number of pedagogical frameworks center teachers' role as cultural brokers as a pathway for fostering positive student outcomes (Grant & Sleeter, 2006; Villegas & Lucas, 2002). I will briefly introduce three such frameworks: multicultural education, culturally responsive teaching, and social justice teaching.

Defining multicultural education (ME) as a concept, an educational reform movement, and a process, Banks (2019) describes ME as having guidelines or dimensions that include content integration, knowledge construction, prejudice reduction,

empowerment through school culture, and equity pedagogy (p. 19). Other pedagogical frameworks rooted in social justice, such as culturally responsive teaching (CRT) (Gay, 2000) and social justice teaching (SJT) (Cochran-Smith, 2010), highlight the importance of “explicit knowledge about cultural diversity” (Gay, 2002, p. 107) and recognition of the range of value systems that different groups ascribe to. Evidence also suggests that socioeconomic and racial/ethnic similarity between students and teachers can impact teacher performance and, later, student performance (Ahmad & Boser, 2014; Bennett, 2003; Gershenson et al., 2016). Bristol and Martin-Fernandez (2019) have reviewed the positive effects on test scores and students’ self-efficacy of racial/ethnic pairing for Black and Latinx teacher–student dyads.

Though these are rich frameworks, all have limitations with respect to the process of guiding teachers toward practice norms that lead to equitable opportunities for learning. These frameworks offer guidelines that work as goals or axes of practice, yet fail to clarify or reflect teachers’ level of alignment with the framework’s dimensions or principles is explicitly offered. Furthermore, despite acknowledgement that teachers’ racial/ethnic and cultural backgrounds are important determinants with respect to the enactment of equitable opportunities for learning approaches (Banks, 2019; Cochran-Smith, 1995), models through which teachers’ personal characteristics interact with their implementation of these approaches are not explicit, either.

Banks (1993) advocates for preservice teachers to have opportunities to “investigate and determine how cultural assumptions, frames of references, perspectives, and the biases within a discipline influence the ways knowledge is constructed” (p. 10). In their review of considerations for the design of courses for teacher candidates, Adams

and Welsch (1993) discuss why teacher candidates must become acquainted with a “sociological view of race” (p. 72) to underpin appropriate meta-awareness about their race/ethnicity and practice.

To describe the ways individuals identify with a racial group, scholars sometimes use race and ethnicity interchangeably with the term “racial identity” (Cross & Cross, 2008; Helms & Carter, 1991). Ethnic identity is a complex construct that refers to the way one feels as a member of societal groups that may have a common national origin, language, culture, and race (Phinney, 2000). While different, national identity may coexist with ethnic identity, since strong feelings of national pride can exist with strong ties to one’s racial group (Der-Karabetian, 1980). Cultural identity, on the other hand, may include racial and ethnic identity, as well as common beliefs, ideologies, traditions, and languages shared with a group (Hoare, 2006; Roland, 2006; Trimble et al., 2003 as cited in Singleton, 2009). I adopt a definition of race and ethnicity as socially constructed categories in this dissertation. Similar to Quintana and McKown (2008), I take the position that race and ethnicity are more similar than different in most people’s lived experiences – in agreement with Cross and Cross’ (2008) view that racial identity is more accurately labeled racial-ethnic-cultural. Focusing on the similarities in ethnoracial identity development and classification in this research does not mean to underscore the important acknowledgement that the real-world consequences of identifying under any non-White category are equivalent (Bean, 2018).

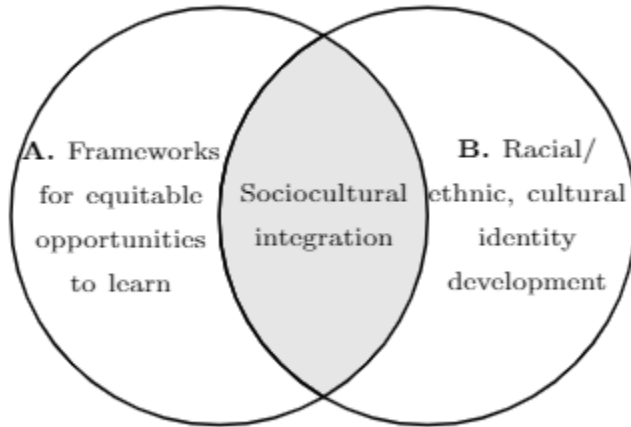
The application of racial/ethnic identity development theory in the field of counseling offers some insight into the role of teachers’ identity development and students’ cognitive and non-cognitive outcomes. As one example, Helms (1984) provides

a framework that explains that a counselor with identity development status self-awareness, which itself requires a high identity development level, can support their client's identity development if the counselor is at a higher identity development level than the client. If we transpose this model to the K-12 setting, each teacher's racial/ethnic identity development level impacts aspects of their practice such as their ability to promote self-esteem and other non-cognitive outcomes or impact student performance due to their establishment of higher/lower standards for certain groups. Positive messaging about one's racial/ethnic group is particularly important for identity development. Quintana and McKown (2008) suggest that the most critical age for identity development corresponds to primary school, which means teachers and peers in primary school settings play an especially important role in the racial/ethnic identity development process.

To address some of these gaps and opportunities with respect to advancing equitable opportunities for learning, this dissertation introduces a nuanced definition of sociocultural integration that recognizes individual educators' positionality in their practice as both a key factor and a potential context for growth. Working at the intersection of sociocultural integration and racial identity development theory puts the focus away from ethnoracial or cultural group affiliation. Instead, the interest lies in the interaction between teachers' racial/ethnic and cultural affiliation, their awareness of it as they move through the levels of racial/ethnic identity development, and their daily practices with respect to SCI. Figure 1.1 provides a graphical representation of the conceptual frameworks that inform this work.

Figure 1.1

Conceptual Model for Definition of Sociocultural Integration



The letter *A* in Figure 1.1 represents educational frameworks that both acknowledge that a teacher’s identity has a role in their classroom practice and propose tools to promote sociocultural integration. Depending on the framework, the tools proposed may be at the individual action, pedagogical, or school level. In Chapter 2, I review the three frameworks I introduced earlier in this chapter: multicultural education (Sleeter & Grant, 1987), culturally responsive teaching (Gay, 2000; Ladson-Billings, 1995; Villegas & Lucas, 2002), and social justice teaching (Cochran-Smith, 1999, 2010; Cochran-Smith et al., 2016). The letter *B* in Figure 1.1 represents the racial/ethnic and cultural identity models I studied to understand identity development progression. In Chapter 2, I summarize and contrast key racial/ethnic and cultural identity development models such as Cross’ Nigrescence theory (1971), Helms’ Black and White identity development models (1984, 1990), Cross and Cross’ racial-ethnic-cultural identity model (2008), and Bennett’s model of Intercultural sensitivity (1993). Finally, the area in grey

in Figure 1.1 is the intersection between these bodies of research, the intersection from which my definition and operationalization of SCI stems.

Using the conceptual framework illustrated in Figure 1.1, I define SCI as knowledge and acceptance of our identity, attaching value to the identity of others, and engaging with others in equitable relationships. This definition is teacher-centered and embedded in a model that considers the promotion of sociocultural integration as a continuum with different levels of endorsement. The operationalization of this new definition is what I use to develop the Sociocultural Integration Scenario Scale for Teachers (SCI-T).

Purpose of the Study

The objective of this research is twofold. First, I propose a revised definition of sociocultural integration that is teacher-centered and explicitly incorporates teachers' racial/ethnic identity development in the evolution of their actions to support SCI. Using this model, I posit that an assessment of teachers' SCI level at a given school can provide a picture of SCI at the organizational level. Teachers' level of SCI and schools' level of SCI are related in the same way teachers' promotion of positive relationships is related to school climate research in this line of inquiry (Thapa et al., 2013). This dimension of positive relationships is what I argue SCI puts under the microscope by accounting for how the racial, cultural, and linguistic backgrounds of teachers impact their behavior. Second, I operationalize this definition and construct a scale for measuring various degrees of sociocultural integration using innovative "lived experiences" scenario items. I argue that the likelihood and effectiveness of initiatives promoting SCI in the K-12 environment can be increased through use of a measurement scale that tracks efforts and

provides actionable feedback about how to move toward higher levels of performance. For this purpose, I designed the Sociocultural Integration Scenario Scale for Teachers (SCI-T) and explored its validity in this dissertation, which is guided by the following research questions:

1. Can teachers' contributions to sociocultural integration at their school be measured in a valid and reliable way using Rasch and Guttman's scenario scale development methodology?
2. Are there differences in teachers' levels of sociocultural integration, as measured by the SCI Teacher's scale (SCI-T) according to teachers' racial/ethnic identification or other psychosocial characteristics?

The SCI-T is well suited to supplement pedagogical frameworks focused on creating equitable opportunities to learn. The SCI-T assesses teachers' promotion of sociocultural integration in their everyday interactions. My adoption of this view is consistent with Lipsky's (1980) model of the public servant's role in public policy implementation. Teachers are "street-level bureaucrats" (Taylor, 2007), who must use their expertise to cater to the needs of their clients (students and their families) following sometimes loosely written policy guidelines in contexts of limited supervision. Teachers' characteristics, beliefs, knowledge, and attitudes all play a role in how they implement policies in their classroom and school. To make an impact through SCI, teachers' racial/ethnic identity level must allow them to guide others in healthy identity development, and they must embrace an antiracist view of their role as an educator. This latter perspective aligns with Kendi's (2019) view that an antiracist lens is required to address the outcomes of historically racist policies. It is my belief that educational policy

that emphasizes sociocultural integration has the potential to reduce inequalities in opportunity/achievement and bring a healing-centered reckoning to K-12 environments.

In addition, the SCI-T represents a departure from existing scales to measure sociocultural integration and related constructs. As I will present in Chapter 2, most such scales were designed under the Classical Test Theory paradigm (CTT), which focuses on a total score that meets various traditional validity and reliability standards (DeVellis, 2003). The CTT approach, however, often produces scores that lack consistent interpretability across samples and settings. On the other hand, the SCI-T was designed using the Rasch/Guttman scale (RGS) methodology (Ludlow et al., 2020), which relies on Item Response Theory to generate scores that provide interpretable descriptions of individuals along a hierarchical progression of sociocultural integration descriptions. Unlike traditional CTT-based scales that marry short-stemmed items with Likert-type responses, the SCI-T uses scenarios that are complex representations of the theoretical levels of SCI. More informative than a numeric total score, the rich descriptions that accompany an individual's score provide context for that individual's actions in relation to SCI at their level of enactment. Having a clear depiction of what one's level of SCI "looks like" may increase the usability of SCI-T scores and allow for connections between scores on the instrument, professional development, classroom and counseling materials, and other interventions.

The work I present in this dissertation was divided into three main stages: conceptual framework and construct definition; scenario development; and data collection, which included qualitative feedback and survey responses. The steps included in each stage align with the RGS methodology for scenario scale development (Ludlow et

al., 2020) and are described in Chapter 3. This process yields a quantitative assessment of a “hard to measure” construct that is authentic in the sense that it captures teachers’ everyday actions alignment with promotion of SCI, instead of professional standards or pedagogical guidelines that equate performance with frequency of behavior. Over many years studying evaluation and measurement, I have learned that quantophobia, the obsession with turning qualitative observations into quantitative measures, has long been a driver for evidence-based policy in education and other social areas. As a result of my work, sociocultural integration will have a better chance of entering policy discussions, given that any school issue with an operational definition and “thermometer” can be monitored and addressed. SCI-T scores and descriptions of the levels of SCI they represent could be used to potentiate professional development, classroom and counseling materials, and other interventions targeting the opportunity/achievement gap in diverse schools.

Significance of the Study

We cannot address persistent gaps in opportunities to learn or claim to practice inclusive education by passive acquiescence. Educators who have aversion to and/or blinders regarding racial/ethnic, cultural, and linguistic diversity need renewed commitment and creative approaches to promote equity. Practitioners and researchers are rising to the challenge. Trailblazing administrators and teachers have come up with programs and resources to introduce the topics of diversity and inclusion in the classroom. Noteworthy examples include the array of classroom and school community resources created by Learning for Justice, formerly Teaching Tolerance (Southern Poverty Law Center, 2021), and the new curriculum piloted by Fieldston Lower School

in New York, which includes student groups discussing bias and imperatives for a diverse teaching force (Miller, 2015).

This work can contribute to discussions of school climate and opportunity/achievement equity for racial, ethnic, and linguistic minorities in three ways. The first relates to the crucial step of giving sociocultural integration an operational definition. Bounding the definition in a meaningful way provides grounds for measurement. To my knowledge, this work represents the first attempt to bring together literature from different fields that describe the relationships between majority and minority groups according to their race/ethnicity, culture, or linguistic background – and apply that scholarship in the school setting. This contribution will allow scholarly discussions to move toward the refinement of a measurement scale for a critical, overarching problem and incorporate the scrutiny and value of the sociocultural integration lens into efforts to address and overcome opportunity/achievement gap patterns.

The second contribution relates to instrumentalization. My research suggests that publicly available and currently adopted instruments to measure school climate and cultural awareness in schools focus on implicit expressions of sociocultural integration (e.g., knowledge, beliefs, and attitudes) and rely on CTT. SCI-T, on the other hand, focuses on actions and behaviors aligned with SCI, rather than implicit expressions. Because items in the SCI-T scale rely on the Rasch/Guttman scenario methodology, an individual's SCI-T scores do not depend on the sample or the items used in the administration of that specific survey. Instead, scores paint a picture of various enactments of SCI expected at different levels of sociocultural integration (“high,”

“moderate,” and “low”), connecting scores invariably to performance stages and allowing for actionable feedback.

The political climate has set the stage for a reinvigoration of diversity as a public policy theme, and the third contribution of this study relates to the advancement of such considerations. Now is an appropriate time to discuss the relationship between diversity and school climate. Not long ago, school climate reform experienced a renewed push from the Department of Education with the launch of resources in the forms of instruments and guidelines for data collection and interpretation (U.S. Department of Education, 2016). Furthermore, recent racial reckonings in and beyond the United States have put historical inequalities, inequities, and systemic racism under the microscope. Work and tools that support policy options in the educational arena have the potential to impact future generations in direct and important ways, offering a path toward society-wide healing and well-being.

I believe this research can and will serve as fuel for educators, practitioners, and academics alike to engage in discussions about school climate that openly address racial/ethnic, cultural, and linguistic diversity by providing a consensual starting point. Even if there is substantial criticism of this work, those reflections will help move the conversation toward productive endeavors (e.g., better definitions or more authentic scenario items) that put us on the path to the widespread promotion of sociocultural integration.

Organization of the Dissertation

This dissertation is organized into five chapters. The objective of this first chapter was to provide an overview of the dissertation and introduce the conceptual framework

used to define sociocultural integration for its operationalization. Chapter 2 is a literature review highlighting existing scholarship that justifies the use of a conceptual framework that includes racial/ethnic identity development models to theorize how teachers implement pedagogical frameworks to promote equitable opportunities to learn. In addition, sociocultural integration is proposed as an implicit or explicit goal of many of these frameworks. Chapter 3 describes RGS methodology and contextualizes its steps for SCI-T scale development, before results from the pilot data collection are presented. Chapter 4 presents the results from a revised version of the SCI-T, including scenario and teacher-level statistics using the Rasch model. Chapter 5 summarizes various lessons from the pilot and the main study, highlights key findings, and explores the implications of the study, as well as questions for future research.

CHAPTER 2 LITERATURE REVIEW

In Chapter 1, I introduced the conceptual framework that informs this dissertation. This framework states that my definition and operationalization of sociocultural integration (SCI) incorporates two key elements: existing definitions and educational frameworks (ME, CRT, and SJT) that implicitly or explicitly include SCI in their goals, and explicit inclusion of teachers' racial/ethnic and cultural identities as an aspect of their ability to promote SCI. In line with these elements, in this chapter I introduce the construct of interest and argue that SCI is an important feature of any educational policy that strives to foster equitable opportunities to learn. The second section of this chapter is dedicated to the link between teachers' racial/ethnic and cultural identities and how they enact SCI in the K-12 setting. I also review some of the leading racial/ethnic and cultural identity models to bring into focus various developmental progressions scholars have put forth for individuals. Building from there, I present models for more than one racial/ethnic and cultural group to extract common features that inform the proposed progression for teachers' identity development in the context of SCI-aligned practices.

In the third section of this chapter, I articulate connections between the promotion of sociocultural integration, racial/ethnic and cultural identity development, and current educational frameworks focused on eliminating inequities in opportunities to learn. While some such frameworks acknowledge teachers' racial/ethnic and cultural identities as important, this factor is not an axis typically included when discussing practice—and I posit that failure to explicitly discuss how a teacher's worldview and default lenses can impact their practice and reinforce differences in implementation. While implementation differences might not cause harm, it is worthwhile to address a key factor that can help

support the implementation of guidelines, standards, and practices in favor of equity with respect to opportunities to learn.

Lastly, I introduce as key aspects for teachers' promotion of SCI their self-identification (i.e., their level of racial/ethnic and cultural identity development), their promotion of healthy identity development in students, and their agency in enforcing equitable relationships. Existing scales designed to measure sociocultural integration or related constructs are also briefly introduced. The objective of this section is twofold: provide evidence that the theoretical frameworks of the educational practice paradigms reviewed fail to explicitly account for how teachers' racial/ethnic and cultural identities interact with their practice in different settings, and justify the use of a new methodology for the development of a scale to measure SCI. Based on this justification, I use the conceptual framework to close with a revised definition of sociocultural integration and situate my use of Rasch/Guttman Scenario scale development methodology to design the Sociocultural Integration – Teacher Scale (SCI-T).

Sociocultural Integration: A Construct of Interest for Equitable Opportunities to Learn

The determinants of academic achievement were among the most prolific research foci and questions in 20th century educational research. More recently, the role of non-cognitive factors in achievement (i.e., unrelated to assessments of ability or performance) has become the main line of inquiry. The contributions of self-esteem, academic self-concept, and racial/ethnic identity have been studied in K-12 contexts to understand trends for sub-groups of the student population according to their racial/ethnic, linguistic, and cultural backgrounds (Awad, 2007; Fisher et al., 2020; Hsieh et al., 2019; Mitchum,

1989; Taggart, 2018). Sociocultural integration is a construct related to these non-cognitive factors known to influence achievement: self-esteem, racial/ethnic identity, and self-efficacy.

The construct of SCI has been explored extensively in the literature on bilingual education. One of its most recognized scholars, M. E. Brisk, defines sociocultural integration as “the ability to function in the larger society as well as in the heritage community” (Brisk, 2006, p. 10). This definition implies that *socialization* and relationships within and outside one’s *heritage community* are important to the promotion of sociocultural integration. Scarr and Salapatek (1973) define socialization as:

The major process in the [general systems theory of society] to account for the *time* dimension; for society to continue across generations requires that older members influence younger ones to carry on in similar ways. Socialization is the process by which the young learn to become acceptable members of a society. (p. 1)

Consistent with this definition is the notion that home and schools are the institutions wherein and whereby individuals are socialized, and therefore critical axes to promote equal opportunities for all citizens (Kearney, 2019). As Potts (2003) clearly articulates: "In a society stratified by race, class, and gender, schools are by no means politically neutral. Schools have been described as major socializing mechanisms that help maintain existing hierarchical relationships of power and privilege” (p. 174).

To make sense of how to foster equitable opportunities to learn through sociocultural integration, I propose a deeper dive into Brisk’s (2006) definition, which implies that SCI requires a dual process within the individual that requires both self-

awareness around one's racial/ethnic and cultural identity (as implied by affiliation to a heritage community), and awareness of racial/ethnic and cultural diversity via the call to "function in society." The former is a process whereby individuals procure a sense of their unique ethnic/racial and cultural identity in relation to experiences connected to racial socialization, ethnic socialization, cultural transmission, socialization environment, race-related messages, cultural parenting, etc. (Bowman & Howard, 1985; Broman et al., 1988; Marshall, 1995; McAdoo, 1985; Phinney & Rotheram, 1986).

Hughes (2003) defines racial socialization as a set of attitudes and behaviors that transmit worldviews about race and ethnicity to children, with the goal of conveying messages to children of color that bolster early socioemotional functioning through promotion of a positive racial identity. While racial socialization is often thought to be particularly relevant to non-White families because systemic racism, social stratification and negative group stereotypes vastly complicate child rearing (Hughes, 2003), *all* families contribute to processes that shape children's understandings of, and attitudes toward, their own and others' racial/ethnic and cultural groups. Schools, via staff and peers, also influence racial/ethnic and cultural group realizations and attitudes (Aldana & Byrd, 2015).

The idea, then, is to convey messages to children of color that bolster child socioemotional functioning through promotion of a positive racial identity and the provision of resources to help children cope with life stressors, particularly racism. Brown and Krishnakumar's (2007) definition of racial socialization goes beyond the explicit and messages regarding intergroup protocol and relationships. The authors state that promoting cross-racial relationships and talking about coping with racism and

discrimination are well within the realm of the construct. Hill (2001) stresses that messages about race and ethnicity, through child rearing and other influences, are shaped largely by the relevant adults' perceptions of the opportunities and challenges children they care for will likely confront in larger society.

The relationship between the promotion of sociocultural integration in the school setting and racial socialization messages about "societal racism, assimilation, biculturalism, spirituality, and egalitarianism" (Bentley et al., 2008, p. 256) is straightforward. Although racial socialization messages are primarily transmitted in the home early in life, the messages received in social environments (from school, peers, and media) quickly compete with and often overshadow the messaging of family members and/or guardians (Adams, 2010). Sociocultural integration is thus relevant as a construct that addresses an individual's "formation of a clear identity and ability to function in both the larger social context and one's own community" (Brisk, 2006, p. 92), and is based on growing awareness of one's racial/ethnic and cultural identity.

The importance of sociocultural integration and its connection to positive racial/ethnic and cultural views is highlighted by Suárez-Orozco and Suárez-Orozco (2001), who find that healthy identity development and the ability to function in both the larger social context and one's own community are essential for children to thrive in different settings such as at home, at school, with peers, and at work. Umaña-Taylor and co-authors (2004) propose that identity development and academic achievement are connected by self-esteem: "if the social climate in which individuals' lives are embedded does not place value on their ethnic group, and individuals experience discrimination or prejudice, they may display lower self-esteem [than their counterparts]" (p. 10). Students

with low self-perception also find it challenging to function and participate in school (García & Ortiz, 2006). In addition, hidden messages about dominant language and culture status convey to students that who they are, their home language, and/or their racial and ethnic background are of little or no value in school or in society at large (Miramontes et al., 2011). Such realities impact language acquisition (for English Learners), academic achievement, educational opportunities, and, consequently, children's lives in school (Dooley, 2005; García-Nevarez et al., 2005; Youngs & Youngs, 2001).

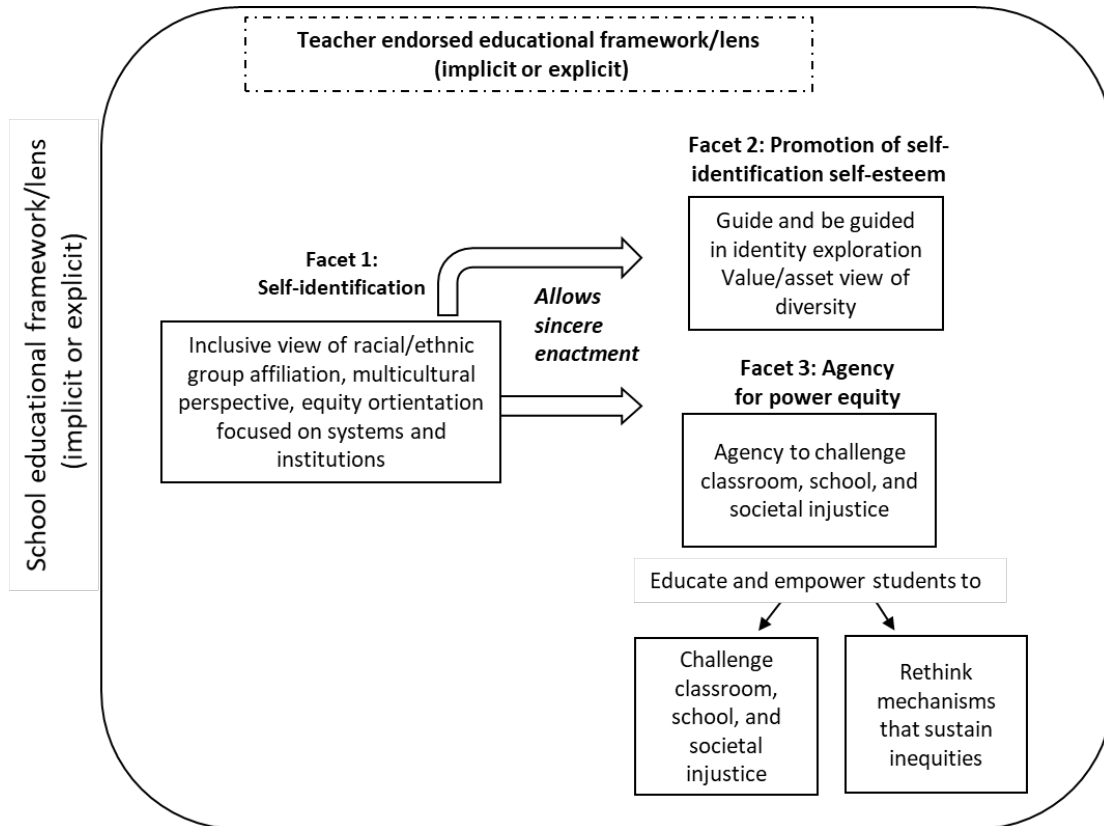
Positive identity has been shown to be particularly key for the cognitive and non-cognitive outcomes of non-White children. For example, Constantine and Blackmon (2002) find that positive identity – operationalized as cultural pride reinforcement and cultural pride – is positively related to important outcomes for Black children, including self-esteem (and academic achievement (Smith et al., 2003). Relatedly, De Jong (2011) has shown that students who speak languages other than English at home experience lower self-esteem and self-efficacy when educators teach them at levels below their cognitive abilities, due to ability misconceptions based on their language acquisition. Zlotick et al. (2019) highlight the impacts of cultural biases such as racism and discrimination on immigrants and ethnic minority youth, which include lower self-esteem and well-being and poorer health outcomes (Acevedo-Polakovich et al., 2014; Hochschild & Weaver, 2007).

Having thus established that the socialization messages individuals receive affect cognitive and non-cognitive outcomes, the next three sections present mechanisms through which behaviors associated with SCI (i.e., socialization messages) play out in the

K-12 context: the teacher’s racial/ethnic and cultural identity, professional guidelines or standards aligned with SCI promotion at the school level, and the teacher’s personal adoption of an antiracist stance that sparks agency and the active pursuit of equitable opportunities to learn. Figure 2.1 illustrates how these principles inform a proposed model for how teachers promote SCI in their daily practice.

Figure 2.1

Proposed Model for Teachers' Promotion of Sociocultural Integration



Self-identification: Racial/Ethnic and Cultural Identity as an Anchor for Teachers' Promotion of Sociocultural Integration

Bronfenbrenner's (1979) bioecological theory provides a framework to understand the relationship between identity (e.g., racial, ethnic, cultural, gender) and sociocultural integration in the K-12 context. In this theory, human development is a process affected by circles of influence. The microsystem comprises the individual's unique characteristics such as gender, socioeconomic status, self-esteem, and support. The mesosystem encompasses the interrelationships between two or more microsystems;

family relationships and school experiences are examples of this circle. The exosystem consists of one or more settings that do not involve direct contact with the individual. Relationships described in the mesosystem are the main setting for racial, ethnic, and cultural identity development as explained by racial socialization. Sociocultural theory offers another contextual link between identity, diversity, and students' success. A child's development occurs in a context, through interactions with adults and other children and participation in "socially shared" activities (Vygotsky, 1934). Both elements will impact learning (Suárez-Orozco & Suárez-Orozco, 2001).

Quiócho and Rios (2000) posit that teachers' identities are "instrumental in teachers' conceptions of diversity" (p. 497). A teacher's identity plays a critical role in their relationships with students and the pedagogical approach followed in their classroom (Duff & Uchida, 1997; Farrell, 2011; Kanno & Stuart, 2011; Menard-Warwick, 2008; Morgan, 2004 as cited in Martin & Strom, 2016). Bennett (1986) posits that teachers' racial/ethnic and cultural identities, which inform their worldviews, influence their behaviors, attitudes, and cognitive frames. Accounting for this dimension of identity, McAllister and Irvine (2000) argue that in order for them to be effective with diverse students, teachers must recognize and understand their own worldviews. Further evidence suggests that only after confronting their own racism and biases can teachers interact effectively with their students, learn about their students' cultures, and perceive the world through diverse cultural lenses (Banks, 1994; Gillette & Boyle-Baise, 1996; Nieto & Rolón, 1995; Sleeter, 1992; Villegas, 1991). Importantly, given increasingly diverse school and societal populations: "[the] racial identity levels of educators

themselves influence how they perceive and interact with children of color” (Carter & Goodwin, 1994, p. 307) .

The K-12 context is also important from the child’s perspective. Erikson’s stages of psychosocial development (1963, 1997) and Green and colleagues’ (2016) environmental identity development model for children suggest that a child’s growing self-concept rests heavily in their ability to form trusting relationships and bonds with educators and caregivers. Peña and Bedore (2010) stresses that during elementary school years having a secure sense of who they are, where they come from, and how their families and communities support their identity is critical for children. Consequently, teachers’ success at promoting healthy racial/ethnic identity development of their students can have positive impacts on their self-esteem, confidence, and motivation – and through these impact learning (Tomlinson, 1996). It is therefore fundamental to review how individuals develop self-concepts related to the racial/ethnic and cultural groups they identify with.

I propose in this dissertation that no model for teachers’ promotion of sociocultural integration is complete without the explicit incorporation of teachers’ racial/ethnic and cultural identity development. As the abovementioned evidence suggests, the extent to which a teacher’s practice aligns with educational frameworks that aim to provide SCI and equitable opportunities to learn is influenced by their race/ethnicity and cultural lens, as well as their racial/ethnic and cultural identity development. The former refers to the teacher’s affiliation or identification with a group, which influences their choice and implementation of racial/ethnic and cultural socialization practices. The latter has to do with the teacher’s self-awareness in relation

their identity. As such, here I will review some of the leading models and theories surrounding identity development in relation to race/ethnicity and culture, to draw commonalities across models and theorize how these stages influence teacher practice.

Racial Identity Models

A series of models have been advanced to explain the ways individuals develop self-concept and their interaction with a group identity. Closely related to racial identity development are Tajfel and Turner's social identity theory (1979), the self-evaluation maintenance model (Tesser, 1988), self-categorization theory (Turner & Onorato, 2014), and Sidanius and Pratto's (1999) social dominance theory. According to Bentley and colleagues (2008), the literature on racial socialization is divided into two strands. Ethnic or cultural socialization reinforces the values and strengths of a specific ethnic group and focuses on preparation for the biases or stereotypes that may be faced due to membership in that group.

Social identity theory (Tajfel & Turner, 1979), social dominance theory (Sidanius & Pratto, 1999), self-categorization theory (Turner & Onorato, 2014), and the self-evaluation maintenance model elucidate the ways individuals see themselves and the groups they interact with (for a review see Singleton, 2009). Here, I will dive into racial identity models directly. Helms (1990) defines racial identity development as the process of developing a sense of group identity based on one's perception of sharing a common racial heritage with a particular group. Helms (1990) recommends looking at the stages of racial identity as a continuum of complexity in one's ability to conceptualize the racial features of the self and those of other group members.

Cross' Nigrescence theory is one of the seminal works on racial identity development (Cross, 1971). In a 1971 study, Cross put forth a conceptual framework for the evolution of African American identity, based on the literature on racial preference. His work contributed to the birth of the multicultural psychology field, alongside cognitive-behavioral, psychodynamic, and humanistic strands (Pedersen, 1999). The work's sociopolitical context included the Civil Rights and Black Power Movements, both of which fought systemic discrimination and racism against African Americans (Cross & Vandiver, 2001). Cross' model describe four stages of African American self-actualization – Preencounter, Encounter, Immersion-Emmersion, and Internalization – associated with eight identities (Cross & Vandiver, 2001). For the Pre-encounter stage, identities include assimilation, miseducation, and racial self-hatred. No identities are attributed to individuals in the Encounter stage. At the Immersion-Emersion stage, identities are anti-White and intense Black involvement. Identities attached to the Internalization stage are nationalist, biculturalist, and multiculturalist. I describe the stages as per Cross' 1991 revised model, which focused more on the overarching theme of each stage rather than the identities associated with each stage (Vandiver et al., 2002).

At the first stage, the Preencounter, the individual feels inferior, lacks self-acceptance, and exhibits low levels of self-actualization (Carter, 1991; Parham & Helms, 1985a, 1985b). Taylor (1986) found that this stage to be linked to feelings of anxiety, depression, and asocial or aggressive social attitudes. At this stage, African Americans tend to place little importance on race but exhibit group orientation centered on being American, or hold negative views about Black people (Black self-hatred). The Encounter stage represents interaction between the individual and external events that makes the

individual reevaluate their reference group orientation. Parham and Helms (1985b) report lower anxiety, higher self-regard, and self-actualization as correlates for this stage. The Immersion-Emersion stage follows, and with it emerge two views about society that Vandivier and colleagues (2001) describe as the “everything Black or Afrocentric is good (Intense Black Involvement)” identity and the “everything White or Eurocentric is evil (Anti-White)” identity (p. 177). In both cases, there is realization of or awakening to a reality and hidden history of Blackness, as well as to injustices committed against Africans and African Americans. Parham and Helms (1985b) suggest that, in counseling contexts, this stage is correlated with high anxiety, low self-regard, and low self-actualization; in addition, an attitude of hostility may be present. The final stage is Internalization. Black self-acceptance is at the center of all internalized identity clusters (Black Nationalism, Biculturalism, and Multiculturalism). Cross (1991) notes that this acceptance of Blackness does not insulate African Americans from depression or change core personality traits. The common thread is that individuals at this stage concentrate their energies on empowering the Black community (Vandiver et al., 2002).

Helms (1984) examines the development of racial consciousness with separate stages for Blacks and Whites, with the purpose of providing theoretical clarity on the influence of participants’ race in counseling. Helms wanted the profession to depart from “one-side cultural analyses in which the ‘problem’ is the minority clients’ cultural adaptations and the ‘solution’ is that the (usually) White counselor either must learn to understand the other culture or else avoid cross-racial counseling situations” (p. 153). In Helms’ contextualization, the interaction between the race and racial identity stages of

counselor and client supports the adoption of these considerations in the K-12 environment.

Helms (1990) reformulated Cross' model to suggest each stage be considered a cognitive template that individuals use to organize racial information. Helms proposed White racial identity development as occurring through six stages in which individuals move from a colorblind view of race to a less racist perspective. The six stages are Contact, Disintegration, Reintegration, Pseudoindependence, Immersion-Emersion, and Autonomy. A basic assumption of Helms' White identity model is that this group has been and continues to be beneficiaries of racism, even if they are unaware of their inheritance from belonging to the dominant culture (Helms, 1984). The Contact stage is the least cognitively mature and characterized by ignorance or a naïve attitude toward the historical roots and sociopolitical implications of racial affiliation. The Disintegration stage is characterized by an incipient awareness of "race-related moral dilemmas" (p. 242). At Reintegration, individuals idealize White culture at a conscious and unconscious level. This stage much like Immersion-Emersion as described in the previously introduced Black identity models. Both include the idealization of one's group and denigration or hatred of the other group (Whites in the case of Black identity models, and any other group for Helms' White identity model). The fourth stage, Pseudoindependence, entails a "White savior" attitude that situates the issues of other racial groups as emanating from their lack of guidance on "how to be more like Whites" (p. 242). Individuals arrive at an "intellectualization" of race-related issues. A psychology concept, intellectualization is a defense mechanism where individuals remove themselves emotionally from a stressful situation, activating reasoning to block confrontation and

stress. Individuals in the Immersion-Emersion stage try to incorporate a nonracist view of White identity and are proactive about teaching other Whites to adopt this view. Finally, there is Autonomy, the most cognitively mature stage in the White identity model. Individuals who reach this stage have gone beyond just internalizing a nonracist White worldview. They actively avoid capitalizing on the benefits of racism and are more cognizant of the historical and current experiences non-Whites live with and endure in the United States.

Helms (1984) also examines the potential outcomes in the counseling practice by counselor-client race dyads (Black-Black, Black-White, and White-Black). The exploration of White clients' interaction with Black counselors is something the author considers understudied – either because of a perception that Whites have no racial preferences or because the proportion of Black counselors is so low that the pairing occurs very seldom. With respect to Black dyads, Helms poses that Black clients in the Preencounter stage might feel insulted when assigned to a Black counselor and express open or covert hostility. A counselor at this stage may adhere to the dominant society's racial stereotypes and “behave in a manner toward the client that confirms these stereotypes” (p. 157). An Encounter-stage Black client may feel insecure and seek approval from their Black counselor. A Black counselor in the same stage will also be seeking approval from their Black client and will likely avoid adopting strategies that may upset the client. The Immersion-Emersion stage has a dual pull reflected in the client's attitude. They may be angry and hostile and require their counselor to pass “covert and overt tests before they are permitted any psychological closeness” (p. 157). The counselor, on the other hand, will be on the lookout for behaviors that are

inconsistent with advancing Black culture, possibly attempting reeducation. The client-counselor interaction may be suboptimal in the communication techniques and care strategies adopted. Finally, clients in the internalization stage will be aware and unafraid to tackle any racial issues, but their issues will be centered not around their Blackness but on becoming the best version of themselves. The Internalizing counselor recognizes that acceptance of one's race is an important part of the actualization process but does not regard race through a deficit or asset mindset.

Helms also adapts Cross' model for White client-counselor dyads. Helms explains that when the counselor's and client's races differ, there are two possible types of relationships. In parallel dyads, the client and counselor share similar attitudes about Blacks and Whites. In crossed dyads, the client and counselor are in stages where their attitudes around Blacks and Whites are opposed. Helms proposes for cross-racial dyads that the counselor cannot move the client further than the counselor has personally come. "To the extent that racial issues are an important concern in the counseling process, regressive relationships are likely to end ... because the counselor is unable to enter the client's frame of reference" (p. 159). For Helms, the racial consciousness model functions like any cognitive developmental model: a person perceives the world (or segments of the world) and reacts and interacts according to their perceptions. Consequently, counselors cannot change their clients' stages of racial consciousness. However, Helms proposes that it is not a helpless situation because training in culture-specific skills and/or exposure to collaborators and peers from different races for clients and counselors alike can lead to positive client-counselor relationship outcomes. This

theoretical conclusion can inform the developmental approach to SCI with regard to teacher-student relationships.

Cultural Identity Models

One reason to go beyond Black and White racial identity models is the common misperception that all individuals from marginalized groups may already be bicultural and have the skills needed to interact effectively with people from diverse backgrounds. Bennett (1993), for instance, argues that marginalized people "may understand and even respect differences with which they are familiar, but they may be unable to recognize or use this sensitivity as part of a generalized skill in adapting to cultural differences" (p. 56). McAllister and Irvine (2000) offer the example of an African American teacher who is fluent in Spanish and feels a cultural connection with their Hispanic students but is insensitive to their Vietnamese students.

The racial-ethnic-cultural identity development model is proposed by Cross and Cross (2008) as a way of interpreting the intersection between racial, ethnic, and cultural identity across the lifespan. Consistent with the previously described ecological conceptualization of development, this model places identity development in the context of human development stages. In early childhood, according to the authors, personal identity is developed. Late adolescence and adulthood subsequently bring the culmination of group identity development.

Cultural identity and racial identity models share features across developmental stages. For example, in Bennett's model of Intercultural sensitivity (1993) the first stages relate to obliviousness or denial of diversity, much like Helms' White Racial Identity Model (1984). Bennett makes a case that this stage is mostly relevant to dominant groups,

“a luxury” only they can afford, whereas “the oppressed” are always aware of their “status.” Denial of diversity is likelier to occur through physical isolation in homogenous communities or neighborhoods.

Bennett describes a stage not unlike Helms’ Disintegration state – Defense – which can lead to defensive postures and adoption of a position of superiority that can occur with or without the denigration of other groups. The Disintegration stage is what happens when the individual becomes aware of the repercussions of race affiliation on a personal level. There is internal struggle between White and Black culture for African Americans, first a visceral reaction to oppression and toward the society upholding this oppressive status quo, and then separation of one’s role in the preservation of oppressive systems. For Helms, Reintegration is characterized by “idealization of everything perceived to be White and denigration of anything thought to be Black.” For Bennett, though, the last phase of the Defense stage coincides with Cross’ third stage of Immersion-Emersion. Banks’ (1994) “Ethnic psychological captivity” presents a similar description of individuals internalizing negative stereotypes and beliefs around their group. The implications of this turbulent stage for non-dominant groups (i.e., anyone who is not European-American) include the pressure to disavow their cultural roots to be truly American (Bennett, 1993, p. 40).

Banks’ (1994) Typology of Ethnic Identity is contextualized for education and presents a model that can be used with individuals of any ethnic or racial group. Unlike the White and Black racial consciousness models presented above, which are formal developmental progressions, Banks’ typology is an “ideal-type construct” (1994). Banks proposes six stages that individuals can move through, albeit in non-monotonically

increasing fashion: ethnic psychological captivity, ethnic encapsulation, ethnic identity clarification, biethnicity, multiethnicity and reflective nationalism, and globalism and global competency.

Other culturally-specific models for certain racial/ethnic groups extend beyond African Americans and Whites, and argue for non-linearity. For example, Horse's (2005) American Indian Identity Development perspectives does not expect individuals to progress linearly through discrete stages, describing instead five influences of the Native American consciousness. The first is the extent to which the individual is grounded in cultural identity that includes their Native American language, rituals, and overall culture. The second is the "validity of one's American Indian genealogy" (p. 65). Third is the alignment of the individual's worldview with the traditional American Indian philosophy. The individual's self-concept as an American Indian and enrollment in a tribe represent the final influences.

Gallego and Ferdman's Latino Identity Orientations (2007) outlines a six-lens trajectory through which Latinx individuals view their identity. The first stage, "White Identified", is similar to Black identity models. At this stage, the individual identifies with cultural aspects and feels affinity with Whites. At the second stage, "Undifferentiated", one adheres to the dominant culture and views the individual, rather than race, as being at the core of sociopolitical realities. Then, in the stages of "Latino as other" and "Subgroup", the individual identifies as "generic" Latinx or with multiple Latino subgroups. The "Latino Integrated" and "Latino Identified" stages include acceptance of the Latinx self-identification and understanding of racial constructs; "Latino Identified" adds empowerment and agency to challenge and promote policies.

Kim's Asian American Identity Development Model (1981) has five stages. Unlike other models, the first stage, "Ethnic Awareness", is described in the context of a specific age range, with children around three years of age expected to be receiving ethnic identity cues from their households. Similar to the lower stages of the Black and Latinx identity models, Kim's "White Identification" stage references a drop in self-esteem and negative ethnic identity attitudes due to socialization outside the home. The third stage, "Social Political Consciousness", is characterized by understanding of oppression as a sociological concept, which leads the individual to stop their identification with Whites. The "Redirection" stage is much like "Immersion-Emersion"; there is buoyant pride of one's own culture. This stage is sometimes accompanied by resentment or hatred of White privilege. The highest form of Asian American identity is the "Incorporation" stage. As with the "Multiculturalist identity" in Cross' model, there is respect for other cultures and feeling of being settled in one's identity. For Kim, this stage also comes with a release from feelings against White culture, to a neutral stance.

To incorporate developmental identity stages into a model for teachers' promotion of sociocultural integration while bypassing assumptions about the teacher's race/ethnicity or culture, I use three "broad" stages common across the models introduced in this section. In a review of the four most widely adopted identity development models (Cross' Nigrescence Model of African American Identity, Helms' Black and White Racial Identity Development Models, Banks' Typology of Ethnicity, and Bennett's Developmental Model of Intercultural Sensitivity), I find the following features noteworthy. These models are conceived as process-oriented cognitive models in which individuals move through a set of stages, with higher levels of identity development

associated with deeper understandings of the complexity of group and individual identification, and higher levels of abstraction in conceptualizing race/ethnicity and culture.

The second feature is best described by Mezirow (1978) and related to the bottom, middle, and top theoretical levels. Individuals move from “a self-centered state,” due to lack of exposure or an ethnocentric view, to identification with society and eventual affiliation with “the larger global community, improving their ability to place their identities or those of others within an increasingly larger perspective” (p. 106). As people mature or move through the models they become increasingly “inclusive, discriminating and more integrative of experience” (p. 106). For a teacher model of SCI promotion, I propose that the higher the teacher’s self-identification, the more receptive the teacher is to SCI-related guidelines or standards, and the more likely the teacher is to harbor the capacities (e.g., perspective-taking and reflection skills) and sensitivity to promote SCI and equitable opportunities to learn.

A final input for my teacher model of SCI promotion is Helms’ notion that the counselor’s own level of identity consciousness anchors how much they can guide their client in that aspect. Helms (1984) goes as far as detailing the outcomes for the client-counselor relationship by levels of racial identity development: parallel when the pair is matched in their level and crossed regressive or progressive when the counselor is at a lower or higher stage than the client, respectively. Because my model is concerned with teachers only, not teacher-student relationships, I only assume that some behaviors to promote SCI will be in the realm of a lower racial/ethnic and cultural identity, while others will be displayed only when the teacher is at a higher level of self-identification.

Table 2.1

Stages of Self-Identification for Teacher Promotion of SCI

Low	Moderate	High
Color-blind perspective, aligned with lowest stages of racial/ethnic and cultural identity development models (e.g., Contact, Pre-encounter)	Heightened awareness and group identity exploration, aligned with middle stages of racial/ethnic and cultural identity development models (e.g., Immersion-Emersion)	Acceptance of self, holding multiculturalist view aligned with highest stages of racial/ethnic and cultural identity development models (e.g., Autonomy)

Frameworks Promoting Healthy Identity Development in Students and Teacher Agency to Pursue Equity

There are numerous mechanisms through which the teacher-student relationship can lead to better cognitive outcomes, as well as higher social and emotional development. For instance, Peguero and Bondy (2011) pose that students' relationships with teachers impact students' self-esteem, cognition, motivation, affect, behavior, and social skills. In this section, I argue that a teacher's ability to confront the complex enactment of culturally relevant, empathetic, and equitable practice is influenced by their racial/ethnic and cultural identity development level, captured by self-identification. Rather than segmenting teacher practices by racial/ethnic or cultural groups, identity development as individuals is more closely aligned with equitable principles and the promotion of sociocultural integration.

In Bristol and Martin-Fernandez's review (2019), the authors highlight how a racial/ethnic match between teacher and student reinforces student self-worth for students of color. In the case of Latinx students, the perception of shared lived experiences and value for linguistic diversity increases students' feelings of school connectedness and

aspiration. This is due, at least to some extent, to what Quijano and Rios (2000) describe as a connection between the personal and professional identities of teachers of color, as well as the use of Spanish language (Reese et al., 2014 as cited in Bristol & Martin-Fernandez, 2019), which helps establish empathy with students. Latinx immigrant students whose teachers care about and are active learners of their complex lives have been shown to have improved schooling experiences (Conchas, 2001). Affective aspects such as care are also salient in recent research about Black students' academic and socioemotional success. Studies show that when Black students are paired with Black teachers, the latter are able to care and motivate their students more effectively than White teachers, due to Black teachers' ability to engage in culturally-responsive ways and serve as role models (Bristol & Martin-Fernandez, 2019).

Educational frameworks have identified dimensions that align with these findings. Positive school climate is one such championed education policy. While school climate does not have a consensus definition, it has four dimensions scholars do agree on: order, safety, and discipline; teaching and learning supports; personal and social relationships; and school connectedness (Thapa et al., 2013; Voight et al., 2013). Research around school climate has been viewed as a viable policy response that serves all students *and* targets underserved student populations (Centers for Disease Control and Prevention, 2009; National School Climate Center, 2007). The National School Climate Center (2007) states that “a sustainable, positive school climate fosters youth development and learning necessary for a productive, contributive, and satisfying life in a democratic society” (p. 4). Related research suggests that behavioral problems are less frequent in

schools where students perceive structure, fair discipline practices, and positive student-teacher relationships (Gregory & Cornell, 2009; Selman et al., 2010).

Sociocultural integration would be closest represented by school climate's dimension of relationships, as the research on ethnoracial student-teacher pairs suggests, although it is also relevant for the dimensions of teaching, and learning and family outreach (Baez-Cruz & Scanlan, 2016). In its guidelines for the relationships dimension, the National School Climate Center (2020) includes respect for individual differences and respect for school tolerance norms. In their review of the research on school climate perceptions according to students' race/ethnicity, Thapa and colleagues (2013) found that race is itself a significant factor in explaining varying perceptions of school climate, and students' self-declared needs in their relationships with teachers vary by racial/ethnic or cultural affiliation. This is further evidence of the complexity of teachers' implementation of educational guidelines for the promotion of SCI and equitable opportunities to learn.

If the relationship dimension guidelines seem broad or superficial, the professional standards for teachers of English learners (ELs) presented by the National Council for the Accreditation of Teacher Education are quite specific – and I argue these should not be limited to the Teachers of English as a Second Language subset of the teacher population. Standard 2, for example, speaks to the teacher candidate's understanding of the relationship between cultural group affiliation, individual identity, school achievement, and language acquisition. This is highly aligned with the teacher model of SCI promotion that internalizes identity development rather than looking across ethnoracial, cultural or linguistic group affiliation. In addition, as with the agency required by higher levels in the racial/ethnic identity models, Standard 5.b. speaks about

the teacher candidate's ability to form partnerships to advocate for ELs (Lucas & Villegas, 2013). Martin and Strom (2016) summarize teacher identity characteristics associated with those who teach ELs, including awareness of their own sociocultural-linguistic identity, sociopolitical consciousness, and willingness to assume responsibility for ELs learning and positioning, and advocacy for ELs. These orientations of linguistically responsive teachers align with equitable opportunities to learn and the promotion of sociocultural integration.

Lee (1997) defines multiculturalism as the ideal state in which people's culture, language, heritage, and humanity are fully valued. Multiculturalism is usually the highest level in the racial/ethnic and cultural identity development models reviewed in the previous section. It makes sense that a pedagogical approach following this ideal stage has grown in popularity and adoption in the past two decades. The modern definition and principles of multicultural education have been the work of Banks (1994, 2019) and Nieto (1992, 1994, 1999). Banks (2019) defines multicultural education as a “reform movement designed to make some major changes in the education of students” (p. 1), with key goals including “help individuals gain greater self-understanding by viewing themselves from the perspectives of other cultures” ... “provide students with cultural and ethnic alternatives [in the curriculum]” ... and “provide all students with skills, attitudes, and knowledge needed to function within their ethnic culture, the mainstream culture, and within and across other ethnic cultures” (p. 2).

This final goal is much like Brisk's (2006) definition of sociocultural integration, providing another indication of the closeness of the concepts and objectives across literatures that can enhance the measurement effort. Nieto (1994, 1999) adds detailed

classroom and school accounts and discusses how aligned everyday actions can be “possible” versus “ideal” in the promotion of the principles of multicultural education. The principles of multicultural education are aligned with the teacher model of SCI promotion that I propose, and this educational framework is more widely adopted than the professional standards for EL teachers. Scales used to measure teacher practice aligned with such educational frameworks are further reviewed later in this chapter – and used to motivate this study.

A social justice education is “centered in democracy and the freedom to exercise one’s full humanity” (Belle, 2019). Cochran-Smith (2010) proposes a framework for social justice teaching and teacher education that strives to challenge systemic social inequities “by attending to the dual dimensions of recognition of social groups and (re)distribution of goods” (Chang, 2017). In this framework, teachers’ self-identification and awareness of the sociocultural and sociopsychological context of themselves and their students is key. Without it, they cannot be critical about the process of knowledge construction, as it has been defined in the pedagogical tradition, or fully comprehend its limitations (Cochran-Smith, 2010). Cochran-Smith also calls for giving teachers who understand that the nature of their work is political and advocate for equity the title of agents of change. Chang (2017) summarizes the guiding principles for instructional approaches under the social justice education paradigm:

- a) Developing caring relationship with students;
- b) recognizing cultural experiences students bring to the classroom as a recourse for the design of curriculum and instruction that are rich, relevant, culturally responsive to

students; and c) making the discussion of equity/inequity and respect/disrespect an explicit part of classroom learning. (p. 16)

Social justice teaching shares with Teaching English as a Second Language standards an emphasis on teacher agency and advocacy. Both frameworks rely on teachers' racial/ethnic and cultural identity development as an enabling factor to adopt equity-oriented practices. With teacher self-identification as a building block, it is easy to also hold a view of the profession as political. Agency for equity promotion is an important feature for a model of teacher enactment of sociocultural integration. In fact, Guardiola (2014) summarizes how key educational stakeholders recognize the importance of teachers proactivity in challenging discrimination and oppression. The National Education Association, for example, the largest employee organization in the United States representing educators, has a Code of Ethics that calls on members to refrain from cultural discrimination.

More recently, culturally responsive pedagogy has been associated with both promoting healthy identity development in students and agency for equity. Culturally responsive pedagogy “engages identities and identity issues across all groups and communities: gender, ability/disability, religion, sexual orientation, socioeconomic status, race, ethnicity, language and nationality” (Southern Poverty Law Center, 2021). Student agency is a key feature of culturally responsive pedagogy. Indeed, according to Ladson-Billings (2009), “culturally relevant teaching is about questioning (and preparing students to question) the structural inequality, the racism, and the injustice that exists in society” (p. 140). While Ladson-Billings proposes a vision of a culturally relevant school that includes specific features focused on African American children and culture, for the

purpose of sociocultural integration we can extrapolate to other groups, as well. Ladson-Billings describes the first key school characteristic as flexibility to determine instructional paths, because local, state, and federal mandates sometimes suffocate the adaptability teachers require to serve diverse students. The second is “accurate and fair representation of African American culture in the school curriculum” (p. 151). The third element, which has come up already in different conceptual contexts, is agency to help students “understand the world as is and equip them to change it for the better” (p. 152).

School districts such as New York City Public Schools District 15 have adopted culturally responsive education as a means of promoting equitable opportunities to learn and addressing biases and inequities in the public-school education system (WXY Studio, 2018). This pedagogical framework connects curricula and pedagogical practices to students’ experiences, histories, and cultures, fostering positive academic, racial, and cultural identities. The ultimate objective is that culturally responsive pedagogy will empower students to create diverse connections and be agents of social change who connect across cultures, race/ethnicities, and languages. A classroom that follows culturally responsive pedagogy makes students feel seen, valued, cared for and respected as their full selves.

Although sociocultural integration is a goal for many educational frameworks, the guidelines to promote SCI are not followed because of a crucial gap: according to Miramontes and colleagues (2011), culture is a foreign concept for many prospective teachers. De Jong (2011) reminds us that many aspects that seem commonplace in the K-12 school experience are actually manifestations of low SCI (e.g., the overrepresentation of minority students, particularly African Americans and bilingual learners, in learning

disability classes). On the other hand, high levels of SCI require school wide initiatives such as parent involvement practices, mission statement creation, and textbook selection to reflect explicit commitment to the inclusion and representation of diverse voices.

De Jong (2011) stresses that both schoolwide and classroom-centered initiatives must align to promote SCI. At the classroom level, teachers could examine, for example, whether students “have opportunities to represent and explore multiple identities and how their practices include their students’ lived experiences in meaningful ways” (p. 176). The standards and guidelines presented by these frameworks implicitly or explicitly include sociocultural integration in their dimensions. However, the decisions and actions required by teachers to address diversity in their students racial/ethnic and cultural identities and their needs (since groups are not monolithic) hold an important assumption. Namely, teachers must be in a high level of racial/ethnic and cultural identity development, and capable of guiding students in their own development – all while catering to their educational needs.

In this review of racial/ethnic and cultural identity models, as well as that of the pedagogical frameworks that promote equitable learning, I identify three key aspects for teacher promotion of SCI: teachers’ self-identification (i.e., their level of racial/ethnic and cultural identity development, which acts as an anchor for other behaviors aligned with SCI promotion), their promotion of healthy identity development in their students, and their agency in enforcing equitable relationships. In the next section, I contrast my view with a selection of existing measures of sociocultural integration and related constructs, drawing weaknesses and strengths in these works, and to motivate the creation of the SCI-T.

Measurement of Racial/Ethnic and Cultural Identity Development, Healthy Identity Promotion, and Agency for Equity

The instrument review process was guided by the key aspects included in my model for teacher promotion of SCI: racial/ethnic and cultural identity development, healthy identity promotion, and agency for equity. My review of quantitative instruments included many surveys; this section will highlight the most influential in terms of the relevance of identified enactments. López's (2017a) Survey of Critical Awareness and Byrd's (2017) school racial climate survey are two examples of recent work in the field of education targeting competencies related to asset-based pedagogy and the relationships dimension of school climate – taking race into account – in K-12, respectively. The first is administered to teachers and the second targets students.

Closely related to sociocultural integration, López's (2017a) "critical awareness" "includes the understanding of the historical context of historically marginalized students; the discrepancy between what is typically validated as knowledge in classrooms and the challenges to those assumptions; and the ways the curriculum in schools serves to replicate the power structure in society" (p. 3). This definition encompasses teachers' identity awareness and affirmation of that of their students, but is missing the aspect of SCI's definition linked to fighting for equitable relationships. Lopez's measure of teachers' critical awareness and asset-based pedagogy is based on Pohan and Aguilar's (2001) Beliefs About Diversity scale. Those authors re-classified the original items to match the four dimensions of asset-based pedagogy (critical awareness, cultural content integration, language, and cultural knowledge), as well as beliefs associated with teacher expectancy and formative assessment. The response option was a 5-point Likert scale

from strongly agree to strongly disagree. In contrast to the objectives of this study, that survey focuses on beliefs rather than actions or teaching practices. Because Lopez's paper does not show the full reclassification of the items, I used Pohan and Aguilar's scale for reference.

Byrd's (2017) School Climate for Diversity – Secondary (SCD-S) Scale follows a two-dimensional framework that includes two domains: intergroup interactions and school racial socialization over ten different subscales. The subscales for stereotyping and school socialization (cultural, critical, mainstream, and colorblind) describe situations related to sociocultural integration in the K-12 setting. These subscales are on a 5-point Likert response scale ranging from not at all true to completely true. The survey has been validated on two independent nationwide samples. Again, though, the items do not reflect the ideas of agency and practice for equitable relationships.

A less recent example comes from D'Andrea and colleagues' (1991) Multicultural Awareness, Knowledge, and Skills Survey Teacher Form (MAKSS Form-T), which follows Atkinson and colleagues' (1979) model of multicultural competencies. This survey is a self-reported teacher questionnaire on three dimensions: multicultural awareness, multicultural knowledge, and multicultural skills. Response options include a 4-point Likert scale from "Very Limited" to "Very Good." While separated theoretically, there is overlap between the awareness and knowledge dimensions. In addition, the MAKSS Form-T focuses on beliefs and knowledge, while the Sociocultural Integration Scale- Teachers focuses on practices and actions. Finally, no items reflect the agency aspects of teachers, who advocate for culturally, linguistically, and ethnically diverse students or peers.

An exception to the neglect of the agency for equitable relationships aspect of SCI, Enterline et al.'s (2008) Learning to Teach for Social Justice-Beliefs scale is grounded on the principle of social activism by teachers on behalf of their students, and calls upon teachers to challenge structural inequalities at school and disrupt teaching processes that limit their students. The scale consists of 12 5-point Likert scale items from "Strongly agree" to "Strongly disagree," including a neutral option (3 = Uncertain), with reported Cronbach's alpha above 0.7. This instrument is also an exception in that it was developed using Rasch measurement principles.

Finally, Siwatu's (2007) measure of Culturally Responsive Teaching includes a self-efficacy (CRTSE) scale and an outcome expectancy (CROTE) scale, consistent with Bandura et al.'s (1977) definition of self-efficacy as a two-dimensional process that requires "the development of a strong sense of efficacy to put acquired skills to use" (p. 1087). Tested initially on pre-service teachers, the scales reflect a battery of Culturally Responsive Teaching Competencies (Siwatu, 2005) rated on a confidence scale ranging from 0 (not confident at all) to 100 (completely confident). Reported reliabilities were above 0.9 for both scales. Like other surveys targeting teachers, CRTSE and CROTE scales focus on what can be called predictors of teacher behaviors (disposition, awareness, and knowledge), rather than end products (i.e., practices and actions).

A champion of school climate research and policy, the U.S. Department of Education's ED School Climate Surveys website (EDSCLS) offers a suite of survey instruments developed for schools, districts, and states by NCES. The EDSCLS follows a three-dimensional model of school climate (fewer than Zullig et al.'s [2010] five dimensions) that includes engagement, safety, and environment, and groups 13 related

topics, of which only one (Cultural and Linguistic Competence) is related to SCI. The measure includes three instruments for students, instructional staff, and non-instructional and principals. The instructional staff instrument includes 82 items corresponding to the three dimensions that use a 4-point Likert agreement scale from “Strongly Agree” to “Strongly Disagree.” Reliabilities in their pilot study were at 0.9 for each of the dimensions (Ye & Wang, 2017). Of the 82 items in this survey, only 4 relate to SCI; 2 are about students (respect for all students’ cultural beliefs and practices and school rules applied equally to all) and 2 are about teachers (being teased or picked on about their race/ethnicity, or about their culture or religion).

Other school climate instruments sanctioned by the Department of Education’s National Center on Safe Supportive Learning Environments that include dimensions relevant to SCI include Zullig and colleagues’ (2010) School Climate Measure. It has eight subscales measuring the dimensions: positive student-teacher relationships (nine items), school connectedness (six items), academic support (six items), order and discipline (seven items), school physical environment (four items), school social environment (two items), perceived exclusion/privilege (three items), and academic satisfaction (two items). All items use a 5-point response scale from “Strongly Disagree” to “Strongly Agree.” The 39-item School Climate Measure was validated with reported reliability above 0.8 in a study including only four of the eight subscales (Zullig et al., 2014). While at least three of the nine items in the student-teacher relationships factor relate to SCI, as do all three of the exclusion/privilege items, their wording shies away from situating race/ethnicity, culture, or language as playing a role and fails to link any specific teacher actions to SCI.

The California Healthy Kids Survey is another climate instrument that includes at least one dimension related to SCI (self-awareness and school connectedness). The teacher module has 139 items in total (using different response options), including about 12 items related to SCI, of which more than half explicitly mention race/ethnicity, culture, or language as playing a role in interactions. Furlong and colleagues (2011) found reliability above 0.8 and concurrent validity ($r = .44$ to $.55$) across 18 sociocultural groups.

Other instruments aim to capture constructs related to SCI outside the school context. For example, the National Center for Cultural Competence's (Goode, 2010) Cultural and Linguistic Competence Family Organization Assessment Instrument aims to assess cultural and linguistic competency of family organizations serving youths. The World View, Peer to Peer, and Advocacy subscales illustrate institutional practices aligned with SCI using a 4-point Likert scale ranging from "Not at All" to "Very Much," and a 5-point scale ranging from "Never" to "Routinely." No peer-reviewed studies use this instrument and report its reliability.

In the field of nursing, the measurement of cultural competencies follows a parallel urgency to education: registered nurses are much less diverse than the populations they serve (Loftin et al., 2013). Rew and colleagues' (2003) Cultural Awareness Scale consists of five subscales, using a 7-point Likert scale ranging from "Strongly agree" to "Strongly disagree," of which cognitive awareness and comfort with interactions are behaviors that could translate into the K-12 context. The reported reliability of the instrument ranges from 0.8-0.9 depending on the tested population. Parallel to the MAKSS Form-T, Perng and Watson's (2012) Nurse Cultural Competence

Scale uses 41 items to measure cultural awareness, knowledge, sensitivity, and skill. With reliabilities ranging from 0.7 to 0.9, the subscales use a 5-point Likert scale ranging from “Strongly agree” to “Strongly disagree.” The Inventory for Assessing the Process of Cultural Competency Among Healthcare Professionals-Revised has been a commonly referenced instrument for gauging achievements in cultural competency because it is a short and easy to administer instrument that has been validated (Kardong-Edgren & Campinha-Bacote, 2008). Measuring five cultural constructs (including cultural desire, awareness, knowledge, and skill), the scale has 25 total items. The Inventory for Assessing the Process of Cultural Competency Among Healthcare Professionals-Revised uses a 4-point Likert scale reflecting the response categories “Strongly agree” to “Strongly disagree,” “Very aware” to “Not aware,” “Very knowledgeable” to “Not knowledgeable,” “Very comfortable to not comfortable,” and “Very involved” to “Not involved” (Ho & Lee, 2007).

In studies of racial socialization, two instruments stand out. Both the Comprehensive Race Socialization Inventory (CRSI) and Scale of Racial Socialization for Adolescents (SORS-A) were designed and tested to serve African American families by surveying adolescents and youth. SORS-A has four factors: spiritual and religious coping, extended family caring, cultural pride reinforcement, and racism awareness teaching; the last two include items with practices that can translate to the SCI Teacher Survey. Lesane-Brown and colleagues’ (2005) CRSI captures critical components of the race socialization process that, according to the authors, “have been absent from existing race socialization inventories” (p. 163) and can be adapted to non-Black racial groups. There is no unique response option for CRSI items, as the instrument includes frequency

of behavior response options in a 5-point Likert scale from “Very Often” to “Never,” “Yes/No” responses for messages received about racial socialization, and open-ended items. Reliability is not reported for the scale, but face validity and predictive validity are established. Both scales include messages around cultural and racial/ethnic pride as items that are adaptable to the K-12 context. While Stevenson’s (1994) SORS-A directly enquires about family interactions, CRSI asks about “receiving information from people,” which can accommodate other adults (such as teachers) in racial socialization roles. However, teachers may be more likely to use these identity formation and pride messages in schools that have an explicit goal for/of sociocultural integration.

Overall, about half of the instruments reviewed use items or draw full subscales from existing scales (e.g., Byrd, 2017; López, 2017; Perng & Watson, 2012), or are not explicit about how items are sampled from the possible universe of items to measure the construct. While some of these studies discuss how they obtained face validity (e.g., Lesane-Brown et al., 2005), authors are not particularly transparent in disclosing the item development process. An additional finding from this review is that regardless of the scales’ population of interest, the focus of the assessment (e.g., beliefs, attitudes, knowledge) is translatable and relevant to the K-12 context. However, the Sociocultural Integration Scale-Teachers will focus on behaviors instead of other aspects that may influence behavior. Thirdly, school climate surveys include too few items (if at all) related to relationships across racial/ethnic, cultural, and linguistic groups, and are thus ill equipped to provide information on SCI at the school. Finally, an important commonality in these scales is that they follow Classical Test Theory principles for instrument design and analysis (excepting Enterline et al., 2008). In the next section, I will discuss

limitations associated with the Classical Test Theory approach and motivate the Rasch/Guttman methodology used in the development of the SCI-T.

Proposed Model for Measuring Teacher Promotion of Sociocultural Integration

In previous sections, I summarized racial/ethnic and cultural identity theories and models; introduced frameworks in education and nursing relevant to the promotion of sociocultural integration; and reviewed some of the scales related to the models, theories, and frameworks I presented. I have stressed that students of color are more likely than their counterparts to encounter instances of discrimination (negative messages around self-identification). These instances require coping strategies for protection and identity-empowering messages that lead to healthy functioning (Scott Jr., 2003). The literature on racial and ethnic socialization has mainly focused on the messages of parents to their children around cultural and racial/ethnic self-esteem. Negative messages around aspects of diversity in a school setting could affect individuals' self-concept and influence their behavior (professional and social).

The scholarship around racial/ethnic and cultural identity development stress that self-identification and appreciation for the identity of others begin with identity formation. Identity formation is influenced by the messages received from our environment: "throughout our lives, we are exposed to verbal, visual, and tacit messages that shape our self-perceptions as well as our interpretations of our place in the world" (Bentley et al., 2008, p. 255). An individual's self-awareness and self-identification are key to determining how much they contribute to sociocultural integration in their social environments. The messages they transmit to others will influence their self-identification

and self-esteem, which is particularly important in environments where individuals are from diverse racial/ethnic, linguistic, and cultural backgrounds.

The first motivation for developing the Sociocultural Integration – Teacher scale (SCI-T) is that my instrument will address both the racial/ethnic and cultural identity of the teacher, as well as the pedagogical practices they adopt. Existing measures for cultural competence and racial/ethnic and cultural identities, separate the context of the respondent (in this case their work as an educator) from their self-identification.

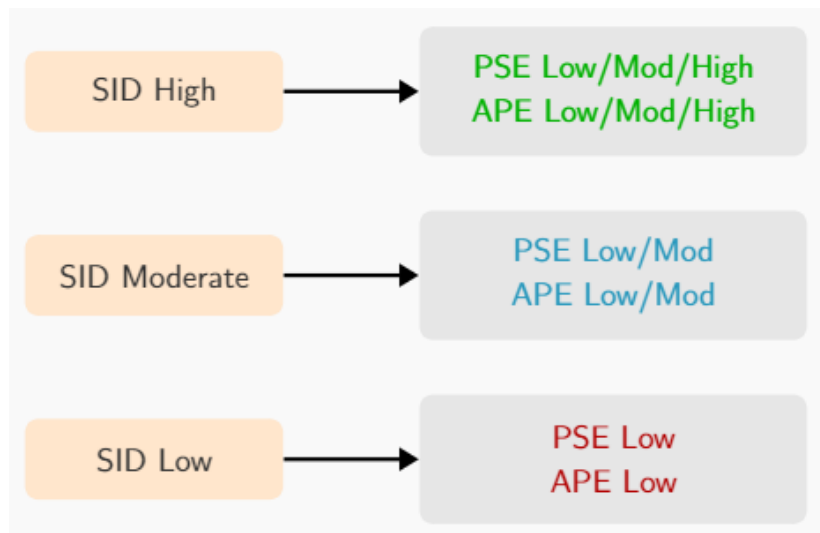
However, evidence summarized in this chapter provides support for bringing both aspects together. Research on student-teacher relationships and widely adopted educational frameworks recognize that teachers' recognition of their racial/ethnic and cultural identity influences their worldview and everyday actions – including their practice as teachers. This is a key and novel takeaway to incorporate into the measurement of teacher practice in line with the promotion of sociocultural integration.

As a result, one contribution of this dissertation is a revised definition of sociocultural integration that capitalizes on the three key aspects that affect teacher practice: their self-identification (i.e., their level of racial/ethnic and cultural identity development, which acts as an anchor for other behaviors aligned with SCI promotion), their promotion of healthy identity development in students, and their agency in enforcing equitable relationships. This definition provides a theoretical basis for a developmental model that links teachers' racial/ethnic and cultural identity to their actions contributing to SCI at the organizational level. I define SCI as knowledge and acceptance of our identity, attaching value to the identity of others, and engaging with

them in equitable relationships. The model that underlies this definition is illustrated in Figure 2.2.

Figure 2.1

Relationship between Facets in Model for Teacher Promotion of SCI



The anchoring property of teachers' racial/ethnic and cultural identity development is illustrated through (a) the arrow going horizontally from that first dimension to the others, and (b) the possible level combinations where a teacher cannot exhibit a higher level on the other dimensions than the one attained in self-identification.

Another important result of the literature review that informs my revised definition of sociocultural integration is that the different constructs, models, and theories around identity development and professional practices that support equity view their target construct as a continuum. Some racial/ethnic and cultural identity development models go as far as instructing readers to expect non-linear growth. The Rasch/Guttman

Scenario scale development methodology (Ludlow et al., 2020) is a measurement framework well-suited for this type of construct. Using this novel approach constitutes an additional motivation for developing the SCI-T.

Three major methodological concerns are addressed by use of the Rasch/Guttman Scenario methodology. First, I argue that the Rasch/Guttman Scenario increases the validity of the scale, compared to the existing instruments reviewed in this chapter. Validity in scales is concerned with how we assign values to a pattern of responses on a scale according to some system of rules (Stevens, 1946). More plainly, validity is related to the extent to which total scores accurately represent what they are intended to measure. Issues of validity can arise when a measurement scale pairs short-stemmed items with Likert-type response options, which happens often when Classical Test Theory instrument development principles are followed (Ludlow et al., 2020, In Press). The second shortcoming is related to measurement invariance. Scale development approaches that follow Classical Test Theory principles use sample-dependent summary statistics to describe item-level and person-level attributes. In consequence, comparisons across scale administrations are futile (Crocker & Algina, 1986) and inference about growth or across different samples are meaningless. This is troublesome for a construct such as sociocultural integration, for which the measurement is interesting in a policy perspective – if it can be tracked.

Another argument to develop the SCI-T is the insertion of transparency and systematic process into the item development method. This is necessary in the measurement of non-cognitive outcomes important in the K-12 context. As questionnaires have become ubiquitous data collection tools, the practice of survey

development and instrument design has become commonplace, with certain instrument design principles applied consistently (Anderson et al., 1998; DeVellis, 2003). At the same time, scales are in danger of yielding “unreliable data of limited validity and utility” (p. 170) without examination of blindly-followed design principles. The high stakes attached to achievement tests drives the emphasis for alignment for items measuring cognitive ability. Since few to none of non-cognitive constructs measured in schools have consequences at the student level, the emphasis in this type of validity is lower. In this sense, the item construction process for the SCI-T is presented in detail, and processes to obtain face validity or reliability across contexts are reported.

The third and most important argument in favor of developing the SCI-T is that items in the Rasch/Guttman Scenario methodology are not short-stemmed statements. Rasch/Guttman Scenario methodology uses scenarios that are complex representations of SCI in the daily lives of teachers. This feature allows the SCI-T to include rich descriptions with individuals’ numerical scores on the scale. The descriptions are more informative and provide a context for teachers’ actions in relation to SCI at their level of enactment. A clear depiction of what one’s level of SCI “looks like” can increase SCI-T result usability, as well as allowing for the linkage between scores on the instrument and professional development, classroom and counseling materials, and other interventions. This is in line with Nieto’s (1994) view of how multicultural education should be adopted by schools:

It is not a monolithic model or one that can develop overnight. The participants in each school need to develop their own vision so that step by step, with incremental changes, schools become more multicultural, and this more

inclusive and more exciting places for learning ... also provide an apprenticeship in democracy and social justice. (p. 8)

In the next chapter, I present the Rasch/Guttman Scenario methodology in more detail. I also describe how I applied it to the development and assessment of the SCI-T.

CHAPTER 3 RESEARCH METHODOLOGY

In this chapter I describe the research methodology I followed for the development of the Sociocultural Integration – Teacher (SCI-T) scale, the Rasch/Guttman Scenario scale framework (RGS), and introduce the psychometric analyses used in Chapter 4 to address my dissertation research questions. This chapter also includes subsections on the target population for the SCI-T and a step-by-step account of how I applied the RGS scale development methodology. Throughout, I will use the term *item* to describe the units of a scale in the Classical Test Theory (CTT) framework and the term *scenario* to describe the same unit in RGS.

Using the Rasch/Guttman Scenario Scale Methodology to Improve Score Usability in the Measurement of Sociocultural Integration

As a reminder, the research questions addressed in this dissertation are as follows:

1. Can teachers' contributions to sociocultural integration at their school be measured in a valid and reliable way using Rasch and Guttman's scenario scale development methodology?
2. Are there differences in teachers' levels of sociocultural integration, as measured by the SCI Teacher's scale (SCI-T) according to teachers' racial/ethnic identification or other psychosocial characteristics?

In Chapter 2, I argued there were three main reasons to use RGS methodology for the development of the SCI-T. Firstly, the thorough process that goes into scenario building increases the content validity of the scale and contributes to the transparency of the item creation process. Scenarios provide a common anchor for response options to a greater extent than often-ambiguous short-stemmed prompts, which can increase

construct validity, providing support for the use of such a scale. Secondly, the CTT methodological framework has some limitations for scale design that make it difficult to compare results across data collection instances and/or over time. Finally, the RGS framework links total scores to a meaningful description of what a construct looks like at different levels of the outcome. This is particularly strategic for the measurement of SCI, given that pairing a construct of importance for K-12 with an instrument with high score usability can lead to widespread adoption. In addition, under the current framework of validity, the suggested use of the SCI-T – examination of schoolwide SCI distribution – allows for individual and group reflection and discussion on growth paths for SCI promotion.

Rasch/Guttman Scenario Scale Development Methodology

The key motivator for using the RGS methodological framework for the development of the Sociocultural Integration – Teacher scale is increasing validity, which centers score interpretation and use as a central concern. The Standards for Educational and Psychological Testing define validity as “the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (American Educational Research Association, 1999, p. 9). The emphasis on score use marks a bit of a departure from the canon of simply assigning numerals to states according to a set of rules by Stevens (1946), perhaps pointing towards Bond and Fox’s (2015) invitation to “refocus some of the time and energy used for data analysis on the prerequisite construction of quality scientific measures” (p. 19).

Focusing on validity, Ludlow et al. (In Press) describe the issues that can arise when a measurement scale pairs short-stemmed items with Likert-type response options,

which happens habitually when CTT instrument development principles are followed. Another critical shortcoming of scale development approaches that follow Classical Test Theory principles is that they use sample-dependent summary statistics to describe item- and person-level attributes (or amount of a specific construct). In consequence, comparisons across scale administrations may be futile, and inferences about growth or across different samples may be meaningless (Crocker & Algina, 1986). To provide context for why these pitfalls occur, I provide a summary of the CTT framework. Then, I dive into the RGS framework and describe how I applied to construct the SCI-T.

A Summary of Classical Test Theory. Methods based on CTT (also known as true score theory) represent the traditional approach to data analysis in the social sciences. CTT's basic principle is that an individual's observed score on an observational tool (X) can be decomposed into his/her true score (T) and an error term (E), which can be expressed mathematically as:

$$(1) \quad X = T + E$$

The most important difference, for the purposes of this dissertation, between Rasch measurement and CTT is Rasch's invariance property. Consider T as a student's ability in algebra; a difficult test may lower their ability measure, X . Yet, if the same student encounters an easy test, their ability measure (X) will be higher, due simply to test design. Consider the exercise of comparison by imagining that I fix the algebra test and administer it to a group of above-average students. The test (and its items) will appear to be easier than when the same test is used with a group of below-average students. Thus, person ability and item difficulty cannot be generalized to other samples of items and persons, respectively, with different distributions (Smith Jr., 2004b).

The CTT framework places emphasis on indicators such as total score reliability, a priority that has survey design implications. For example, items that are replications of the construct (i.e., target the same level and definition) increase reliability and give more precise scoring estimates (Lord & Novick, 1968). As a result, items in the scale end up being highly correlated and with difficulty levels (defined as the proportion of correct responses) around 0.5 for maximum variability. Another downside of having total scores as the indicator of success is that item-level statistics rarely come into play for an iterative process of item quality review, and measures of item fit (i.e., cohesion with the rest of the items) prize items that are replications, leading to low performing item statistics for those that might measure extreme levels of the construct, which increase construct representation. In addition, missing data means indicators of person ability or item difficulty cannot be compared truthfully, forcing the researcher to employ a missing data imputation or estimation method (Smith et al., 2002), which is not desirable or feasible in some cases. Finally, the emphasis on total scores is not just to determine how psychometrically sound these instruments are; the interpretation and use of the instruments are also guided by scale-level indicators. These scores rarely tend to be anchored in a unit that is meaningful or related to performance levels that can guide actionable feedback (Ludlow et al., In Press). For further discussion on the weaknesses of CTT compared to RGS methodology, see Chang (2017), Ludlow and colleagues (2020), and Reynolds (2020).

Scales designed using RGS methodology constitute a viable and attractive alternative to assess constructs that are hierarchical in nature. This approach to scale development draws from *Rasch measurement theory* (Rasch, 1960), and *facet theory's*

item construction techniques (Guttman, 1954, 1959). Items developed using this approach, called *scenarios*, combine rich descriptions as stems that provide a common anchor for response options (Ludlow et al., 2020). This common anchor increases the scale's validity by limiting the respondent's need for interpretation when providing answers and increasing comparability across respondents. The Rasch measurement framework's property of invariance, when the data fit the model, provides consistent outcomes on the scale for more than one response pattern (Wright & Douglas, 1977). This highly attractive feature is referred to as *person and item free* estimation.

Rasch Measurement Theory. Social science researchers capture observations of human behavior using tools that assign numerals to their observations. At best, these numerals can be deemed ordinal indicators unfit for certain data analysis methods (Bond & Fox, 2015). Rasch measurement theory is a framework that includes analytical methods for the analysis of such ordinal data (Iramaneerat et al., 2008). A key aspect of the Rasch measurement model is rooted in the idea of the model being “specifically objective” (Rasch, 1960/80), so that the two sets of parameters of interest, namely person ability and item difficulty, can be estimated separately. As a result, individuals and items can be uniquely ordered according to their ability and difficulty, respectively:

It is further essential that comparisons between individuals become independent of which particular instruments – tests, or items or other stimuli – within the class considered have been used. Symmetrically, it ought to be possible to compare stimuli belonging to the same class – measuring the same thing – independent of which particular individuals within the class considered were instrumental for the comparison. (Rasch, 1960/80, p.vii)

The assumptions that derive from the model's conceptual backbone are few but important: *unidimensionality*, *monotonicity*, and *local independence*. *Unidimensionality* implies that every item on the scale measures the same trait. More formally, the probability of a "correct" (i.e., endorsing a more difficult item) response π_{ia} for individual a can be considered a value that may be estimated (or observed) for item i by presenting the item repeatedly to individual a . As a result, all individuals a could be ordered according to their level of π_{ia} for items i . It follows that an instrument designed following Rasch principles should only include items that order individuals a according to their level of π_{ia} for items i , if the probabilities were known, and that these *probabilities increase monotonically* along the continuum of the attribute (Mokkan & Lewis, 1982). It should be noted that the assumption of unidimensionality does not mean that performance on selected scale items is strictly caused by a single psychological process. According to Iramaneerat and coauthors (2008), it is sufficient if the processes represented by the items are themselves affected by the same underlying process; this is important with respect to the *facetization* process described later in this chapter.

Unidimensionality and monotonicity are related assumptions, as I have already introduced the concept of monotonically increasing. More practically, I propose that sociocultural integration is a unidimensional trait. As such, SCI can be represented like a staircase where individuals with higher levels of SCI are located further up, and the probability of them endorsing the more difficult items measuring SCI goes up the higher they are on the staircase. Linked to this assumption of monotonicity is the concept of *invariance*. In the Rasch paradigm, the measurement system is modeled in a way that the order of individuals according to their level on the attribute being measured and the order

of items according to their difficulties are invariant. As such, a teacher with higher enactment of sociocultural integration should always have a higher probability of getting an SCI-T scale item “right” than a teacher who enacts lower levels of SCI at their school, regardless of the particular set of items they encounter. The same property holds for items: a more difficult item should always have a lower probability of being answered correctly or endorsed than an easier item, regardless of the attribute level of the teacher who answers (Marais & Andrich, 2005; Rasch, 1960/80; Smith Jr., 2004b).

The assumption of *local independence* refers to the individual’s response sequence. The answer provided for item i should not depend on the answer offered on item $i-1$, so that the probability of answering item i correctly does not depend on whether the individual responded correctly to previous items (Rasch, 1960/80). If this assumption goes unmet, alternative orderings of items would lead to different individual ability rankings (Choppin, 1983). It must be noted that the *item free* and *person free* estimation properties of Rasch models hold only if the data fit the model, which can only occur when there is a clearly defined construct evidenced by items mapping onto the progression of the construct according to the items’ theoretical location. In this sense, the model is fixed and the data are chosen to fulfill the model requirements (Marais & Andrich, 2005). A final assumption is that the probabilistic functions that describe the interactions between individuals and items will be only nonintersecting item response functions.

The measurement model, with its assumptions, is governed by a series of principles that ensure that the SCI-T scale is constructed to provide data that fit the Rasch model. Presented in Ludlow and colleagues (2014), these principles are:

unidimensionality and independence, per aforementioned assumptions; variability, a requirement that the items in the SCI-T target all the levels of SCI on that imaginary staircase from very low to very high levels of SCI; spread, which alludes to how close together the items are (they should be spread all along the conceptual staircase); a requirement that the different levels of SCI follow a clear progression; scenarios must have equal discrimination (each should be related in the same way to sociocultural integration); and, finally, a requirement that the theory behind my conceptualization of SCI should be reflected by the data, meaning that the data collected must fit the model and not the other way around.

Using Facet Design to Create Authentic and Representative Items. My review of scales related to sociocultural integration in Chapter 2 revealed that the process of item generation for a survey of scale is a typically a mix of borrowing from existing surveys and scales and adapting items from existing surveys and scales to serve the study purpose. This practice is not exclusive to SCI. Even in scale development papers on a breadth of constructs and areas of inquiry, the process of item generation is given little space, providing little context for weaknesses that may be found when looking at the instrument's psychometric properties.

Chang (2017) suggests that a scale's construct validity is "often an afterthought in the process of instrument development" (p. 28). Furthermore, it is seldom explicit whether authors seek to capture the construct of interest in its *full expression*. Full expression refers to the breadth of representations and their range; for example, the construct's meaningful variation from "low" to "high" and how it is expressed or manifested in different contexts. This is the key design aspect the RGS inclusion of facet

theory (FT) accomplishes: scenarios as representations from a vast universe of sociocultural integration manifestations:

An attitude or opinion – whether simple or complex – is defined in terms of a universe of items or possible observations, rather than in terms of a single item or observation with its attendant artifacts and biases. It is the statistical structure of the universe that is at issue and the attendant psychological implications. We are now learning how to study many infinite universes simultaneously by the use of a smaller sample of observations than is for most surveys done today, and with richer and more objective results. (Guttman, 1954, p. 396)

Developed by Louis Guttman, FT is an approach specifically designed for inquiries into complex multivariate events (Shye, 1981). Facet theory can offer a systematic approach to bridge the observations (i.e., items developed) and the target construct domain of interest, which provides stronger evidence to instrument validity (Canter, 1996; Hackett, 2014, 2016). The rationale for the relationship between facets, the universe of the construct, and items is well captured by Shye (1988): “the concept common to all item ranges to be determines the boundaries of the content universe delineated the items, concepts that item domain (questions determine the internal structure of that universe...)” (p. 3).

It follows that each item on a scale that is measuring a unidimensional construct can be classified under a “domain facet” (Guttman, 1954) that divides the universe of this target construct into substantive parts. Through its analytical approach, smallest space analysis, FT can illustrate a correspondence between a construct partitioned into facets and responses to items that correspond to these facets (Shye, 1988). Facet theory has

design and analysis components. For RGS, we borrow only the design. FT design's purpose is to identify the facets of a construct to aid in the systematic construction of items from the universe of possible items that might define a construct.

Facet theory design employs mapping sentence structures to combine the construct facets or elements into items, enabling the survey instruments based on this systematic procedure to thoroughly assess the entirety of the construct's spectrum. "[A] verbal statement of the domain and of the range of a mapping, including verbal connectives between facets as in ordinary language" (Shye, 1978, p. 413), a mapping sentence consists of "a formal part made up of the facets, and a less formal part comprising the phrases linking the facets together" (Guttman & Greenbaum, 1998, p. 16). We can think of the mapping sentence as FT design's semantic structure. To build SCI-T scenario items using the mapping sentence technique, I had to choose levels of SCI to target in the instrument, as well as the facets that associate with sociocultural integration. The different levels of the target construct's facets are known as *structs*, and a *structuple* is the combination of *structs* to round out a complete mapping sentence (Borg & Shye, 1995; Guttman & Greenbaum, 1998).

Using FT design and the mapping sentence structure to design SCI-T as contemplated in RGS methodology makes the item and instrument design more transparent and replicable, and promotes harmony with Rasch measurement principles. Other scholars recognize the benefit of following this design approach (Canter, 1996; Chang, 2017; Hackett, 2014; Ludlow et al., 2014; Rossi & Anderson, 1982). In their guide to survey development, Rossi et al. (1982) use a facet-theory driven approach to outline a five-step general process that promotes a more streamlined and efficient item-

generation procedure, guided by the objective of placing items into a "meaningful order and format" (p. 202). This process speaks to Rasch principles of monotonicity and spread. Further symbiosis between FT and Rasch measurement principles is provided by Canter (1996, p. 40) when enumerating FT's "major constituents": a formal definition of the variables being studied, hypotheses of some specified relationship between the definition and an aspect of the empirical observations, and a rationale for correspondence between them.

I have established the main methodological features of the Rasch measurement approach and facet theory, and how they have principles that align. In summary, RGS scale development methodology combines Rasch measurement principles and the Rasch analysis model with facet theory design to ensure transparency and item- and scale-level Rasch principle compliance. Now I will introduce the model estimation details corresponding to the RGS methodology.

Using the Rating Scale Model to Estimate Item and Individual Teachers' SCI Levels. Following the assumptions outlined in the previous sections, I employed the Rasch rating scale model (Wright & Masters, 1982) to assess the success of my theoretical model of SCI compared to the empirical data (i.e., teacher responses to the instrument). The estimation of Rasch parameters provides numerical locations for each scenario and teacher along the sociocultural integration continuum. The result from using the Rasch model in combination with the scenarios is that any teacher's score on the scale will be a minimal sufficient statistic to estimate their level on the unobservable trait (Lord & Novick, 1968, p. 429). The beauty in this result is that I should be able to use any combination of scenarios on varying samples of teachers, and each teacher will obtain a

consistent level on the SCI scale, unaffected by the choice of scenario or the other teachers they happen to take the survey with.

Rating scale model (RSM) estimates for the teacher and item levels of SCI are the core focus of the psychometric analysis in this dissertation. Before estimating the RSM, though, I examined descriptive statistics of the item responses and teachers' response strings. RSM estimates were used on a dataset of teacher-level responses where response options are awarded a *score*, with higher values of the score aligned with higher levels of SCI enactment. Descriptive statistics also included the frequency of response categories across scenarios, which allowed me to assess the performance of the rating scale and identify categories with low frequencies, which can be problematic for estimating model parameters. Next, the RSM was estimated. In the Rasch paradigm, person parameters are usually referred to as ability levels. In this case, interpretation of the parameter was the teacher's level of SCI enactment. Scenario items were characterized by their difficulty, which I interpreted as the effort required to act according to one level of SCI compared to the other levels, as well as the category thresholds, which can be thought of as the level of SCI the teacher's actions will represent that take them from one response category level to the next (e.g., the actions through which a teacher goes from behaving in ways moderately aligned with SCI to behaving in ways highly aligned with SCI).

This idea can be formally represented using the Rasch rating scale measurement model for polytomous¹ responses (Andrich, 1978; Wright & Masters, 1982) presented in Equation 1, which is appropriate to use “when the response categories are intended to

¹ I chose the polytomous data model because the response options are not binary but include five categories.

have the same meaning for all items and the same monotonic order of increasing difficulty across all items” (Ludlow et al., 2014, p. 136).

$$\Pr\{x_{ni} = x, |\beta_n, \delta_i, \tau_i\} = \frac{\exp \sum_{k=0}^{x_{ni}} (\beta_n - (\delta_i - \tau_i))}{[1 + \sum_{k=0}^m \exp \sum_{j=0}^k (\beta_n - (\delta_i - \tau_i))]} \quad (1)$$

Where the probability of teacher n endorsing category x on scenario i is related to the difference in the person’s level on the SCI (β_n), δ_i is scenario i ’s difficulty on the SCI scale, and τ_i represents the x^{th} threshold location of the SCI scale from one response category to the more difficult, subsequent category. The empirical estimates of teachers’ levels of SCI, scenario difficulty, and thresholds for the rating scale model were obtained using unconditional maximum likelihood estimation (De Ayala, 2009). Model estimates of teacher levels of SCI and scenario difficulty are in *logits*. This unit of measurement can be defined as natural log of an odds ratio (Cox & Snell, 1989). “Odds ratios tell the relative frequency of an event occurring versus the relative frequency of it not occurring” (Ludlow & Haley, 1995, p. 969). For the purposes of interpreting the results in Chapter 4, the intuition that teachers with high levels of SCI, and scenarios that are more difficult to endorse will have a positive logit; while easy to endorse scenarios, and teachers with low levels of SCI will have negative logits.

Key psychometric analyses include the model estimation to obtain teacher- and scenario-level parameters, the Rasch-Thurstone *variable map*, the scale and scenario *category characteristic curves*, *Rasch Andrich thresholds*, and analysis of the *residual matrix*. Ludlow and colleagues (2020) consider the variable map “one of Rasch measurement’s most useful tools for assessing congruence between one’s a priori

hypothesis about the intended scale structure and the subsequent empirical results” (p. 372). The variable map presents the distribution of scenario difficulty estimates on the right-hand side, and individuals’ estimates of the attribute on the left-hand side. These distributions are joined by a line in the middle, representing the construct’s continuum, which goes vertically from lower to higher levels of the attribute.

This constitutes the empirical representation of the theoretical staircase I have used to describe SCI progression and is the first test of how my theoretical model of SCI enactment compares to teachers’ responses. The variable map gave insight into my second research question regarding potential differences across teachers’ racial/ethnic identification. Comparing variable maps created for each group, I was able to test whether SCI enactment was different, which I hypothesized would not be the case, and to contrast scenario- and teacher-SCI-level distributions. The latter helped shed light on how scenarios functioned across groups (i.e., scenario/enactment rankings).

Category characteristic curves, which are used to check the distribution of response categories (Smith et al., 2002) at both the item and scale level, express the probability of choosing a response category for a given scenario or the scale, as a function of the teacher’s underlying level of SCI. Category characteristic curves allow inspection of the ordering of response options according to the theory, answering: is the probability of the response category in line with the level of SCI? I introduce in Equation 1 the term τ_i , representing the x^{th} threshold between two consecutive response categories. The SCI-T has five response options, as I will describe later in this chapter, and four thresholds to be estimated. In a well-functioning scale, the Rasch-Andrich threshold cumulative probability curves are distinctive for each threshold (i.e., located at

different levels of difficulty) and have similar distributions (i.e., their curves have similar peaks spread out at different levels of difficulty).

Analysis of the residual matrix, which is the matrix of observed responses minus the expected responses for each teacher for all scenarios, is common to test the assumption of unidimensionality. This analysis is conducted using principal components analysis, a “data reduction strategy to condense a large number of interrelated variables into a smaller set of ‘components’ with minimal loss of information” (Baglin, 2014, p. 2). In practice, principal components analysis produces new variables that are linear combinations of the initial scenarios, so my k-dimensional residual matrix is reduced by principal components analysis to a small number of principal components by condensing maximum possible information in the first component. Ideally, no principal components will remain in the matrix residual because the residuals should be uncorrelated. Principal components analysis constructs a covariance matrix containing the spread and orientation of the data (i.e., the initial residual matrix). *Eigenvalues* are the magnitudes of vectors in particular directions from the components resulting from the principal components analysis. In Chapter 4, I extracted all components with a minimum explanatory power and meaningful common variance for the pilot and the final studies. I expected the extracted components not to be meaningful, using an eigenvalue threshold of 3.0, a threshold put forth by O’Connor (2000) as the comparable value for random data. I obtained a graphical representation of residuals in a two-dimensional space, where I expected residuals to be distributed in a random-like fashion.

Using these diagnostic tools, scenarios were revised in an iterative fashion, until they represented a “ladder-like, uniform, continuous hierarchical scenario structure”

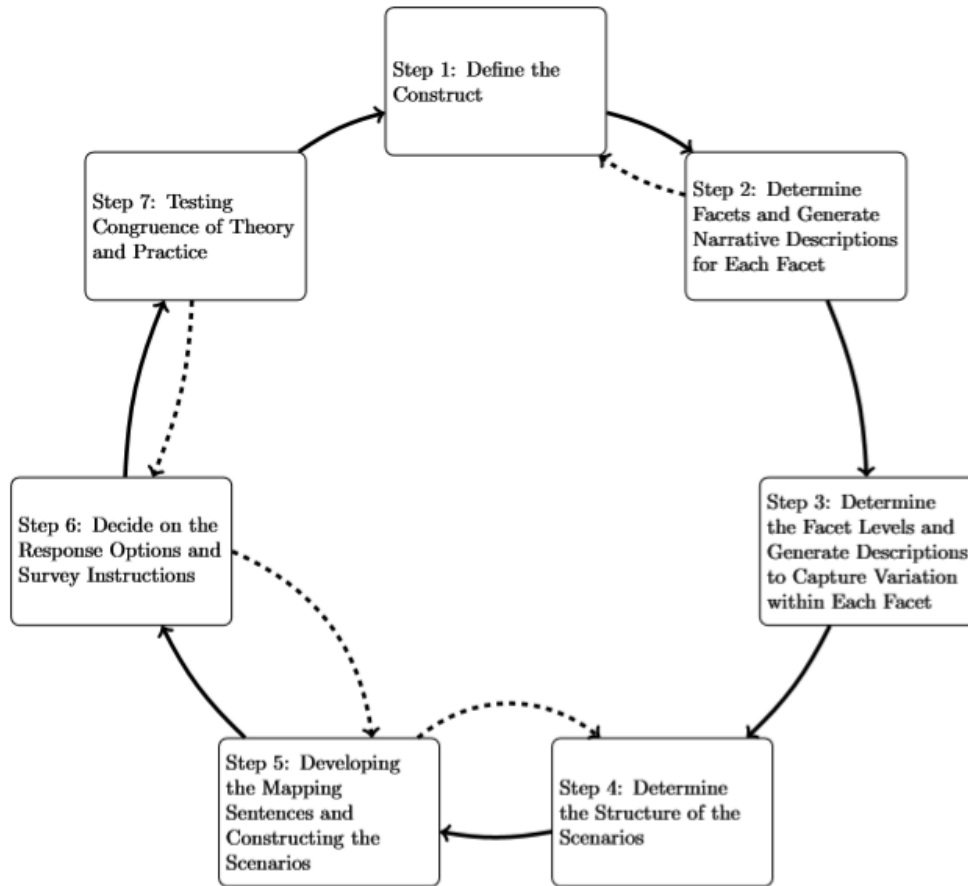
aligned with this dissertation’s definition of sociocultural integration (Ludlow et al., 2019). This is consistent with the RGS methodological framework “trial and error” process of “systematic revisions” to obtain empirical confirmation of the theoretical model (Ludlow et al., 2020, p. 375).

Research Design

Given the rationale for the use of a new measurement framework for sociocultural integration outlined in Chapter 2 and the benefits of the Rasch/Guttman Scenario scales methodology presented already in this chapter, RGS is theoretically suitable for the development of the Sociocultural Integration – Teacher scale. Ludlow and coauthors (2020) present seven steps for the implementation of scale development following RGS methodology: (a) define the construct, (b) *facetize* the construct, (c) determine facet levels, (d) establish the scenario structure, (e) create mapping sentences, (f) develop response options and survey instructions, and (g) test congruence of theory and practice. Figure 3.1 illustrates these steps and shows that the process is not just a one-directional flow. Steps linked with dashed arrows are moments when scenario revisions usually take place, requiring the developer to reflect on feedback, make changes, and re-test.

Figure 3.1

Diagram of the RGS Scenario Development Process



Source: Ludlow et al. (2020).

Applying RGS to sociocultural integration will be part of an increasing trend of employing the RGS approach in educational settings to make sense of patterns such as faculty availability outside of class (Reynolds, 2020), teacher enactments for social justice (Chang, 2017), and parental involvement (Antipkina & Ludlow, 2020). The remainder of Chapter 3 focuses on how each of the steps for RGS scenario development were followed for the development of the SCI-T.

Population of Interest

A key design question concerned the groups of teachers, and inevitably the type of schools, that this instrument will serve. The rationale for developing the SCI-T came from studying school climate at schools within the Two-Way Immersion Network of Catholic Schools (TWIN-CS)². TWIN-CS schools, as well as schools with an international focus (e.g., international baccalaureate), were a purposive sample for the initial stages of SCI-T development because such schools' mission statements tend to explicitly affirm linguistic diversity and cultural appreciation. Furthermore, because these schools are perhaps more racial/ethnic and linguistically diverse in terms of staff and student populations than an average K-12 school in the United States, they represent a key audience for whom this instrument may be of interest. That said, demographic transitions will be impacting nearly all K-12 schools in the United States in the medium term, requiring proactive policy responses to address educational outcome and learning opportunity gaps through the lenses of equality and equity. Looming changes in the makeup of the K-12 student population require antiracist educational policies, which I posit should include ongoing measurement of sociocultural integration. With respect to the population of interest of this study, while I use a purposive sample of schools with dual language programs for the scale design stages (pre-pilot and pilot), the main study includes K-12 teachers from all types of schools.

² TWIN-CS is a national network of elementary schools transforming from monolingual English to two-way immersion models of service delivery. More information can be found on their website: <https://www.bc.edu/bc-web/schools/lynch-school/sites/roche/Programs/twin-cs.html>.

Given their mission, I assumed schools with dual language programs likely had curricular, pedagogical, hiring practices and other structures in place to promote high levels of SCI. Consequently, in the pilot phase I expected scores to cluster on the higher end of the scale and make the results of the pilot truncated from the bottom. Because the SCI-T population of interest includes all types of K-12 schools and teachers, my recruitment efforts were intentionally broader and I expected the levels of SCI from teachers in the main study to be more spread out, and to include teachers at lower SCI levels than teachers in the pilot study.

Step One: Defining Sociocultural Integration

Scale development following the RGS methodology requires intimate knowledge of the construct of interest. As a reminder, my own definition of SCI is knowledge and acceptance of our identity, attaching value to the identity of others, and engaging with others in equitable relationships. Reformulating what sociocultural integration means was a way to bound the construct and update it based on recent research. By bound I mean that teachers and SCI can be related in many ways, with SCI reflected in teachers' attitudes, beliefs, and knowledge. For the purpose of this dissertation, I developed the SCI-T around a definition focused on teachers' actions or enactments. While there are unobservable processes that drive action, teachers' actions are what students react to, so it seemed most sensible to focus on enactments. In addition, the model I propose for teacher enactment of SCI hypothesizes that it is the teacher's racial/ethnic and cultural identity development, not their racial/ethnic/cultural affiliation, that influences their underlying SCI level. The motivation for these elements, and for my proposed definition, are examined in Chapter 2.

After conducting a literature search for the term *sociocultural integration* and related terms identified in Chapter 2, I compiled available definitions and descriptions and used NVivo (QSR International, 1999) to identify common elements and summarized the result in a word cloud (see Appendix 1). One of the implicit assumptions attached to my definition of SCI relates to the subject of the construct. While the goal of this dissertation is to propose a measure of SCI at the school level, the definition is at the individual level. The proposed use of the SCI teacher scores is to create a schoolwide SCI distribution baseline to inform appropriate interventions. This will promote a community lens in the use and interpretation of SCI-T results, which is related to the concept of validity I used to guide this research. A noteworthy consideration is the unit of measurement. As a school level phenomenon, the ideal way to capture SCI would be a 360-degree instrument like those used for school climate. However, to limit the scope of this dissertation and relative convenience of the sample, I focused on teachers' contributions to their school's SCI.

A second assumption pertains to the nature of the construct. I assume SCI is a unidimensional continuum. Furthermore, I propose that the unidimensional SCI is made up of three highly correlated elements or facets. Because these facets are not independent but correlated, to measure SCI I assessed them together as a whole. In this sense, SCI is a unidimensional construct and its measurement well-served by scenarios that account for this complexity. A direct consequence is that individual teachers obtain a unique score on the SCI-T scale that corresponds to a combination of the person levels for all three facets – a description richer and more authentic than separate facet scores that narrowly describe each facet.

I argue that it is possible to identify actions for each facet, and for SCI as a whole, that correspond to high, low, and moderate levels of the construct. It then follows that SCI can be represented using a variable map that graphically represents these levels as a progressive continuum. I make a theory-based assumption of unidimensionality, suggesting that scenarios and individuals can be placed on an empirical “staircase” of sociocultural integration, with those with low SCI at the bottom of the ladder, those with moderate SCI in the middle, and those with high SCI closer to the top.

For construct validation, I sought out expert advice from members of my committee who have experience in K-12 research with racial/ethnic, cultural, and linguistically diverse populations. Additionally, leaders at Boston College’s Roche Center for Catholic Education provided advice and shared their expert opinions. My work at the Roche Center and its Two-Way Immersion Network of Catholic Schools (TWIN-CS) provided motivation and context for this dissertation. Roche Center leaders work with Catholic schools to strengthen their practice through leadership programs, professional education, and research. In this context, construct validation is obtained in a cognitive lab when in-service teachers and administrators complete a test or survey and verbally report their thoughts related to the instrument using both “think-aloud” (concurrent verbal reporting) sessions and interviews with the researcher after each scenario is completed (retrospective verbal reporting) (Levine et al., 2002, p. 1). Findings from the cognitive lab conducted at the TWIN-CS’s 2017 Summer Academy suggested that teachers and administrators recognized SCI as a construct that impacted their schools and everyday practice.

Steps Two and Three: Operationalizing SCI Using Facets, and Determining Levels

With a clear definition of sociocultural integration, I proceeded to facetize the construct, using FT design to frame the facets of SCI. Facets provide “structure given an unstructured universe” (Borg & Shye, 1995, p. 32). In this case, the universe is the construct itself, as well as the sources that “reveal” the construct (i.e., behaviors and/or performances and the tasks or situations that elicit those behaviors). The three facets of SCI I chose correspond to the most prominent descriptors from the construct’s definition, and the more salient features in the literature review: knowledge and acceptance of our own identity, attaching value to the identity of others, and engaging with others in equitable relationships.

For efficiency, I created the following three facet acronyms: self-identification (SID), promotion of identity self-esteem (PSE), and agency for power equity (APE).

I evaluated each facet and sought in the literature related concepts to provide further theoretical grounding and support. Using the definitions of SCI and related constructs described in Step 1, I compiled instruments designed to measure those constructs (see Chapter 2). From these resources, I compiled the behaviors and situations that elicited each facet – SID, PSE, and APE – and rated them according to whether they reflected high, moderate, or low levels of each facet. A byproduct of this process was the construction of a scenario bank that lists the actions/behaviors by facet, their source in the literature, and the associated level of the facet – as judged by me. The scenarios reflect

everyday actions by teachers and administrators that align with low, moderate, or high levels of SCI.³

Before constructing scenarios, based on the scenario bank I named and described each of the three facets, the first being SID. Conceptually, I use SID as the anchoring facet for sociocultural integration. In practice, “anchoring” means to me that individuals are limited in how much their actions support SCI in other facets by their level of racial/ethnic/cultural identity. This anchoring feature is described in Chapter 2, where I discuss the limiting features of the counselor–patient relationship in Helms’ model (1984). SID means to capture the teacher’s lens when interacting with students and other adults at their school. The literature on racial/ethnic and cultural identity reviewed in Chapter 2 provides theoretical support for this facet.

The second facet is PSE, which is related to the self-esteem of students. In Chapter 2, I present evidence that links students’ self-efficacy, socioemotional, and academic outcomes to their self-esteem and a positive racial/ethnic and cultural identity. I also argue for accounting for the role teachers have in their students’ racial/ethnic and cultural socialization, especially for younger students. I hypothesize that teachers’ own SID levels limit the level of promotion of self-esteem they can enact.

The third and last facet, agency for power equity, captures teachers’ antiracist and equity-supporting behavior. Its higher levels test how far teachers [would] go out of their way to support equality at their school to challenge power dynamics, in acknowledgement of their role in preserving systemic inequality. Discussion of equity in

³ I chose to use three discrete levels (low, moderate, and high) for convenience while I obtain proof of concept for the facets. As Ludlow and colleagues (2019) show, I can fill in the gaps to make SCI look more like a continuum in future iterations.

teaching, social justice teaching, and professional standards in support of equitable educational outcomes in Chapter 2 provides theoretical support for this facet.

Steps Four and Five: Formulating Sentence Mapping Structures and Constructing Scenarios

After defining the facets and identifying actions that embody them, I developed a first round of scenarios. Scenarios are composed of a stem that provides a rich description of a hypothetical individual, built using the mapping sentence structure. The description presents the actions of the hypothetical individuals on the three facets, representing a low, moderate, or high level on each. The response options ask the individual to compare themselves to the hypothetical third party, which then reveals the respondent's level on the underlying trait.

Unlike other item types, scenarios provide a common anchor for responses, increasing the validity of the direct comparison of responses across individuals. Likert response options paired with short item stems may lead to responses that are not comparable across individuals, as each person interprets the response options according to personal anchors (Primi et al., 2016). In addition, "Likert-type items tend to exhibit a positive bias in responses that produces ceiling effects and skewed distributions that are not useful for differentiating between people" (Ludlow et al., 2014, p. 130). On the other hand, scenarios can represent the full spectrum of a construct with impressive authenticity, as the construct of interest's variable map corresponds fully to its theoretical description. I decided to use scenarios because my objective was not just to measure, but to design an instrument with high authenticity to capture sociocultural integration in a valid way. I built SCI-T scenarios using a sentence mapping structure technique

described earlier in this chapter. Like other works using RGS, the scenarios I built combined “structs,” which represents an element of the facet, across the three facets of SCI. The following is an example of a mapping sentence:

Facet 1: Self Identification

The teacher speaks (high [with pride/extensively])
(moderate)
(low [uncomfortably/succinctly])

about their roots and upbringing.

The teacher

Facet 2: Promotion of Self-Esteem

(high [encourages])
(moderate)
(low [discourages])

Students to discuss their feelings and fears in the wake of ICE raids.

Facet 3: Agency for Power Equity

Students believe the teacher (high [does everything in his/her power])
(moderate)
(low [does only what the school mandates])

to make immigrant students feel included, valued, and safe.

Extreme Scenarios. I began the scenario development process by creating only extreme-level scenarios for qualitative examination through the qualitative data collection instances described in Step 1. The examination of extremes provides “proof of concept” (Borg & Shye, 1995; Ludlow et al., 2014), while avoiding cases of ambiguity.

Extreme scenarios combine actions that represent the highest (and lowest) level of each facet. The next step was to consider which cases would be ambiguous and/or implausible, given the construct definition. One consideration comes from having defined SID as an anchoring facet. A person’s level on self-identification provides a theoretical limit to their enacted level of PSE and APE. As such, some facet-level combinations are

theoretically implausible⁴ (e.g., a person enacting a low level of self-identification and high level of agency for power equity). Appendix 2 presents the facet/level combination matrix with all plausible combinations marked and implausible combinations blank.

To build an extreme “high SCI” scenario I used three structs: a “high” level self-identification action, one “high” level promotion of self-esteem action, and one “high” level agency for power equity action. I drew the structs from the action scenario bank for each facet on the extreme levels. Consistency in the scenario structure is a key feature in reducing construct-irrelevant variation.⁵ For this reason, structs always start with the name of the hypothetical individual and are separated into sentences.⁶ Below is a revised scenario after feedback on the first version of the SCI-T.

Gloria acts differently when around people whose ethnicity is different from hers, so she tries to fit in by toning down her personality to make others feel comfortable (SID-Low). Gloria gets anxious and experiences frustration when she has conferences with families whose culture is different from her own (PSE-Low). She has never asked the parents if they feel nervous or frustrated when they meet with her (APE-Low).

That scenario was included in the cognitive lab. I discuss cognitive lab results in my discussion of the last step of RGS application.

⁴ I say “theoretically” implausible. While I could have tested whether SID is empirically an anchor facet, that was beyond the scope of this dissertation.

⁵ Construct-irrelevant variation is a term used in testing and psychometrics to refer to extraneous features of an item that affects the individual’s response but is not related to the construct of interest. In consequence, the answers provided for the items should not be interpreted lightly, because their true meaning is masked by what individuals interpreted from the item (given these foreign features) instead of what we asked.

⁶ This structure was my decision based on earlier work by Chang (2017) and Ludlow et al. (2014). However, a simple structure could look different.

Step Six: Selecting Response Options and Survey Instructions

The response format for the scenarios is crucial for obtaining the respondent's level of SCI. Respondents were asked: "In the scenario described, how would you act compared to Gloria?" and provided the following options:

- a) I would act more extreme than Gloria
- b) I would act about the same as Gloria
- c) I would act less extreme than Gloria

In the case of Gloria's scenario, respondents who act "more extreme than Gloria" would be signaling they have higher levels of SCI than Gloria, and thus be in the "moderate" or "high" category. Those who select that they would act exactly as Gloria did would be revealing their level as "low." These response options were not well understood; cognitive lab results suggested they were "clunky." I asked lab participants their perceived scenario difficulty, assigning them a number from 1 to 5, where 1 was the easiest and 5 the most difficult. Most participants reported having difficulty comparing themselves to the teachers in the scenarios using the given response options. In particular, the word "extreme" had negative connotations that biased respondents against the option. In addition, the directionality of the "extremeness" was not well understood.

Other researchers who have used RGS methodology reveal that selecting the response options is one of the more critical decisions in the scenario design stage, and acknowledge that getting it right may require multiple iterations (W.-C. C. Chang, 2017). The response options used by Ludlow and colleagues (2014) are an elegant example; they use a single verb ("engaged") and "more/less than." Table 3.1 summarizes response

options I considered and used for further expert review; the response options I used for the pilot study and the main study are presented thereafter.

Table 3.1

Response Options Considered and Tested for SCI-T

Cognitive Lab Response Options	Alternative #2
a) I would act more extreme than (teacher name)	a) My involvement would be much higher than the teacher
b) I would act about the same as (teacher name)	b) My involvement would be higher than the teacher
c) I would act less extreme than (teacher name)	c) My involvement would be about the same as the teacher
	d) My involvement would be lower than the teacher
	e) My involvement would be much lower than the teacher
Pilot Study Response Options	Alternative #3
a) I would be much more proactive than Teacher X	a) I would be much more embedded than the teacher
b) I would be more proactive than Teacher X	b) I would be more embedded than the teacher
c) I would be about as proactive as Teacher X	c) I would be about as embedded as the teacher
d) I would be less proactive than Teacher X	d) I would be less embedded than the teacher
e) I would be much less proactive than Teacher X	e) I would be much less embedded than the teacher
Main study Response Options	Alternative #4
a) Much more proactive	a) I would be much more active than the teacher
b) More proactive	b) I would be more active than the teacher
c) About the same	c) I would be about as active as the teacher
d) Less proactive	d) I would be less active than the teacher
e) Much less proactive	e) I would be much less active than the teacher
Alternative #1	Alternative #5
a) I would be much more engaged than the teacher	a) I would be much more woke than the teacher
b) I would be more engaged than the teacher	b) I would be more woke than the teacher
c) I would be about as engaged as the teacher	c) I would be about as woke as the teacher
d) I would be less engaged than the teacher	d) I would be less woke than the teacher
e) I would be much less engaged than the teacher	e) I would be much less woke than the teacher

Table 3.1 shows the evolution of the reference teacher in the response options. In the cognitive labs I used teacher names that identified with different racial/ethnic and cultural groups, similar to the famous Bertrand and Mullainathan study (2003). The cognitive lab results suggested that while the names used were meant to be race neutral, as both could be perceived as White or Hispanic names, participants made assumptions that could hypothetically influence their response choice. Names were consequently eliminated. For the pilot study, I moved from three categories to five categories, to increase variability. After feedback at this stage, I also decided to change the hypothetical from “Teacher X” to “the teacher” to avoid connections to African American rights activist Malcolm X. For the main study, references to the hypothetical teacher were removed completely from the response categories and instead moved to the scenario itself, and a prompt before the response options read, “Compared to this teacher I am:”.

RGS scales referenced in Ludlow and colleagues (2020) “ask participants to compare the lived experience captured in a scenario with their assessment of themselves or their perceived circumstances” (p. 371). Given this untraditional item type and response options, instructions for the SCI-T had to be sufficiently clear to avoid any construct-irrelevant variance. Instructions from the pilot and final studies are contrasted in Table 3.2.

Table 3.2*Survey Instructions Tested for SCI-T***Cognitive Lab**

The following items describe a generic teacher and how the teacher acts in school around their peers, their students, or students' families. After carefully reading the description of the teacher, you will be asked to determine how similarly you would act in the situations described. Specifically, at your school, would your actions be more extreme, about the same, or less extreme than the teacher being described? Try the example item and then, slowly, go through the other descriptions and select one response for each item.

Pilot Study

The items in this survey are an innovative kind. You will have to read each one read carefully before providing a response. Each item describes a generic teacher and how this teacher acts at school around peers, students, and students' families. This description has a few sentences. Please consider the scenario holistically. After examining the description of the teacher, you will be asked to determine how your actions on any given day compare to the generic teacher's.

To each scenario respond: how much like Teacher X would you act if you were in this exact situation?

Main Study

The items in this survey are innovative. Each item presents a scenario that describes a teacher's worldview and interactions at school with colleagues, students, and students' families. Some scenarios might seem similar, but each is distinct. After examining the teacher's description, you will be asked to compare your worldview and actions to that teacher's on any given day this school year. Please consider each scenario holistically. There are no right or wrong answers. Read each scenario carefully before providing a response.

Step Seven: Testing Congruence of SCI Theory and Practice

To obtain proof of concept and test scenarios iterations, I conducted three important studies: the cognitive lab I described in Step One, a pilot study, and a main study. Each moment of data collection had a subsequent expert review with two objectives: to discuss results emerging from the data, and discuss potential revision alternatives. In this section, I discuss the scenario revisions resulting from the cognitive lab and pilot study. Suggested revisions emerging from the main study are discussed in Chapter 5. Results of the analysis of responses are presented in Chapter 4.

The cognitive lab with teachers and administrators from the Two-Way Immersion Network of Catholic Schools (TWIN-CS), representing schools in 13 states, provided rich information and insights about scenario authenticity, readability, and validity for a preliminary SCI-T with 17 total scenarios. The 31 teachers and administrators who participated in the cognitive lab provided their scenario responses in paper form. I gathered additional data with voice recorders as participants shared their reactions to scenarios as they read them aloud for the first time, and their impressions regarding the authenticity of the enactments in each scenario.

To maximize feedback, the 17 tested scenarios were divided and administered to one of three participant groups. The first group received five scenarios and the second and third groups each received six scenarios. The scenarios represented only the two extreme levels of SCI (high and low), with each level combining each of the three facets in the target level homogeneously (i.e., SCI High Level = Enactment of High-level SID + Enactment of High-level PSE + Enactment of High-level APE). The distribution of the instruments' design is presented in Table 3.3. I used the following three-point response option: I act more so than (Teacher name), I act about the same as (Teacher name), and I don't act like (Teacher name). The scale range was not maximized to see variability, as the objective of the pre-pilot was to obtain "proof of concept" evidence that sociocultural integration was experienced by pre-pilot participants in the ways described in the scenarios. A more complete table is included in Appendix 3, which includes the scenarios used.

Table 3.3*SCI-T Design for Cognitive Lab*

Group assigned	Scenario	Level	Facet/level combinations
Group 1	SCIT-1.1	High SCI	SID_H + PSE_H + APE_H
	SCIT-1.3	High SCI	SID_H + PSE_H + APE_H
	SCIT-1.5	High SCI	SID_H + PSE_H + APE_H
	SCIT-1.2	Low SCI	SID_L + PSE_L + APE_L
	SCIT-1.4	Low SCI	SID_L + PSE_L + APE_L
Group 2	SCIT-2.1	High SCI	SID_H + PSE_H + APE_H
	SCIT-2.3	High SCI	SID_H + PSE_H + APE_H
	SCIT-2.4	High SCI	SID_H + PSE_H + APE_H
	SCIT-2.5	High SCI	SID_H + PSE_H + APE_H
	SCIT-2.2	Low SCI	SID_L + PSE_L + APE_L
	SCIT-2.6	Low SCI	SID_L + PSE_L + APE_L
Group 3	SCIT-3.3	High SCI	SID_H + PSE_H + APE_H
	SCIT-3.5	High SCI	SID_H + PSE_H + APE_H
	SCIT-3.6	High SCI	SID_H + PSE_H + APE_H
	SCIT-3.1	Low SCI	SID_L + PSE_L + APE_L
	SCIT-3.2	Low SCI	SID_L + PSE_L + APE_L
	SCIT-3.4	Low SCI	SID_L + PSE_L + APE_L

The qualitative data captured by the audio provided more insight into the research objectives of the pre-pilot than the scenario responses. Scenario responses per group ranged from six to eleven, a small sample size for conducting Rasch model analysis. However, classical statistics like average difficulty were calculated. Qualitative data of the “think aloud” exercise revealed overall encouraging feedback. First, there was generalized agreement on SCI as a phenomenon that participants experience in their schools. Regarding authenticity, some participating educators expressed feeling validated by the frank scenarios while others discovered the experiences of their peers through the discussion. Second, the participants referenced the teacher names used in the scenario as revealing something about their identity, in particular their race/ethnicity. While, as mentioned, the names were meant to be race neutral, participants made assumptions that

could hypothetically influence their response choices. Third, the question rating of perceived scenario difficulty captured how participants interacted with the response options. Most participants found it difficult to compare themselves to the teachers in the scenarios with the given response options. In particular, the word “extreme” had negative connotations that appeared to bias respondents against the option and the directionality of the “extremeness” was not well understood.

These pre-pilot results led me to revise the scale in three areas. The more problematic issue, the need for alternative response options, was addressed in the survey pilot. The next improvement had to do with the scenarios themselves, particularly the enactments. I eliminated scenarios that were consistently rated by participants as difficult to answer and reviewed those that received mixed ratings. To fall under the “review” category, feedback on the scenario had to be coded as “inconsistencies.” The inconsistency code was applied when participants noted that the sentences did not “hold as one” (i.e., how they referred to the structs or enactment that corresponded to each facet) or one sentence in the scenario made participants rethink their response because it was “different” from the rest of the scenario. Appendix 4 includes a table with scenario revision options resulting from cognitive lab feedback.

As I introduced in Step Seven, scale development following RGS methodology requires investing in feedback to improve scenario validity. After conducting expert review with the revisions that stemmed from the cognitive lab, I prepared an updated version of the SCI-T to pilot in June 2018 at the Roche Center for Catholic Education’s Summer Academy. However, last-minute scheduling changes did not allow me to

participate in the event in the same manner as the previous year and data collection was not possible.

After gathering further feedback from two presentations at New York University's Faculty First Look program in fall 2018 and spring 2019, I revised the enactments selected for the reviewed SCI-T scenarios. Peers and professors validated the idea of eliminating teacher names, as participants might try to find patterns in the behavior of any given name across scenarios. I was also encouraged to consider whether the enactments that corresponded to each facet/level combination would be likely representations for individuals across race/ethnicity groups (i.e., would the high enactment of self-identification look the same for a Hispanic teacher as it would for an African American teacher or a White teacher?). I addressed these considerations in meetings with members of my dissertation committee and settled on the facet/level enactment combinations presented in Table 3.4.

Table 3.4 should be read horizontally to see within each facet the enactments that correspond to the three levels of measurement: low, moderate, and high. In some facet/level cells, I have more than one enactment. This was my proposed solution to the comments received during Faculty First Look presentations with respect to being inclusive of race/ethnicity in the enactments in certain cells. Read vertically, the table reveals what enactments of the different facets look like within each level of SCI. Appendix 5 shows how the facet/level combinations were implemented for all the available enactments within cells to construct the scenarios tested in the pilot. Appendix 5 also shows the ordinal codes/ scores assigned to the three enactment levels (i.e., High = 3, Moderate = 2, and Low = 1), such that each scenario score is the sum of the facet/level

combination score used for each. For example, the first two scenarios (SCIT-1 and SCIT-2) capturing the high-level enactment have a sum score of 9 that reflects the three facets in their highest level being combined. As a result, if the data fit the model, these two scenarios will be the hardest for participants to endorse (i.e., to say that they are more proactive than Teacher X). Similarly, the last two scenarios (SCIT-13 and SCIT-14) were built combining each of the three facets at their lowest level, with a total score of 3. The underlying hypothesis for these two scenarios is that they will be easy to endorse by teachers. Taking the full instrument, I hypothesized that the 14 scenarios would form a five-cluster structure with the two highest-level scenarios located at top of the variable map, followed by the four high-moderate-level scenarios, then the four moderate-low scenarios in the middle, and the two low-level scenarios at the bottom.

Table 3.4*Facet and Level Combinations Used in Pilot Study Scenarios*

Facets	Facet Level		
	High	Moderate	Low
Self-identification (SID)	SID_H1: (Name) is aware and distressed about how systemic, unearned racial/linguistic advantage and conferred dominance plays out at school and the community.	SID_M1: (Name) perceives racial/or linguistic advantage and conferred dominance in some situations at school and the community. He/she does not consider them systemic.	SID_L1: (Name) does not conceive that racial/or linguistic dominance could affect him/her or his/her students.
	SID_H2: In the presence of representations of different cultures, (Name) has spoken to his/her students about the difference between cultural appreciation and the unacknowledged or inappropriate adoption of cultural features.	SID_M2: In the presence of representations of different cultures, (Name) feels uncomfortable about cultural features being adopted inappropriately.	SID_L2: In the presence of representations of different cultures, (Name) does not know how to assess if any cultural features are being ignored or are adopted inappropriately.
	SID_H3: If a joke charged with bias is spoken in his/her classroom, (Name) immediately walks his/her students through its negative assumptions.	SID_M3: If a joke charged with bias is spoken in his/her classroom, (Name) is uncomfortable walking his/her students through its negative assumptions.	SID_L3: If a joke charged with bias is spoken in his/her classroom, (Name) will change the topic.
Promotion of self-esteem (PSE)	PSE_H1: (Name) actively seeks ways to integrate lessons and materials about current issues affecting Latino people and communities into the curriculum.	PSE_M1: (Name) does not include in his/her lesson plans items that encourage pride in his/her students' cultures, but he/she makes connections in class that serve that purpose.	PSE_L1: (Name) defers the responsibility of encouraging pride in students' cultures to the school.

PSE_H2: When (Name) meets a new student or parent, he/she asks them which language they prefer to engage in conversation. He/she lets them know the languages he/she speaks and the interpretation services available at the school.

PSE_H3: When he/she see students' languages being devalued in school, (Name) speaks up no matter who the offender is.

PSE_M2: When (Name) meets a new student or parent with Spanish surname, he/she immediately addresses them with 'Hola, ¿cómo estás?'

PSE_M3: (Name) is worried and upset when he/she sees students' languages being devalued in school, but (Name) does not speak up.

PSE_L2: When (Name) meets a new student or parent who has brown skin, he/she immediately addresses them with 'Hola, ¿cómo estás?'

Agency for power equity (APE)

APE_H1: (Name) has explained to parents how standardized tests may be biased against immigrant students because of their linguistic or cultural/ethnic background.

APE_H2: After last year's events, (Name) has encouraged immigrant students in his/her class to discuss their feelings and fears around ICE raids.

APE H3: (Name)'s students consider his/her classroom a safe environment where they can be heard. Students feel comfortable using the board to draw things they are trying to describe, or switch languages at any time they need.

APE_M1: (Name) has discussed with peers but not parents how immigrant students might do worse in standardized tests because of their linguistic or cultural/ethnic background.

APE_M2: After last year's events, (Name) has asked fellow teachers whether he/she should talk to students about ICE raids.

APE M3: (Name) wants to advocate for equity in his/her classroom but he/she is just one teacher with too many students in his/her class. Diversifying lessons is difficult for him/her.

APE_L1: (Name) regrets not having sought counsel from peers or supervisors when dealing with linguistic or cultural misunderstandings with students or families.

APE_L2: After last year's events, (Name) defers to the school's principal to discuss anything related to ICE raids. He/she will not engage in a conversation about the topic with students or parents.

Appendix 6 presents the full instrument used in the pilot. In the final instrument, scenarios were manually randomized for two purposes: so they would not appear in an obvious ascending or descending order that might be noticeable by participants, and so scenarios with a facet/level in common would not confuse participants and make them believe the scenarios were the same. Finally, survey instructions were more elaborate than in the pre-pilot, and an example scenario was added at the beginning to avoid start-up effects (Chang, 2017).

In expert review after the pilot study, I was asked to incorporate enactments of SID closer to their theoretical frameworks but applied to the educational context. I conducted a new review of online resources from varied sources (e.g., social media, blogs, and newsletters). In addition, I made sure to include facet representations that resemble actions reported more often for different racial/ethnic and cultural identity affiliations in these new sources and as identified in expert reviews. Appendix 7 includes the full instrument used in the main study. Results of the pilot and final studies are presented in Chapter 4.

CHAPTER 4 RESULTS

This chapter presents the results of the pilot and final administration of the Sociocultural Integration – Teacher scale. For each data collection instance, I discuss the recruitment strategies used, provide descriptive statistics about participants and scenarios, and the results of using the Rasch model to estimate teacher and scenario-level statistics. In order to better understand the rationale behind the results presented in this chapter, I reintroduce the research questions that the study undertakes: (a) can teachers' contributions to sociocultural integration at their school be measured in a valid and reliable way using Rasch and Guttman's scenario scale development methodology?, and (b) are there differences in teachers' levels of sociocultural integration, as measured by the SCI Teacher's scale (SCI-T) according to teachers' racial/ethnic identification or other psychosocial characteristics?

Answering these questions required two instances of quantitative data collection. In order to build an RGS scale with desirable psychometric properties, scenario items had to be constructed and tested. The pilot data analyzed to assess scenario functioning included the responses provided by teachers to the SCI-T pilot version. Accompanying the psychometric and quantitative data, I performed other "tests" using qualitative data. In particular, to establish construct and face validities I conducted cognitive labs and interviews, whose input allowed for scenario revisions to inform the SCI-T main study version. It is worth noting, as described in chapter 3, employing the RGS methodology to develop a scale helps maximize content validity by design. Criterion validity, which establishes the predictive or concurrent validity of a new instrument compared to a well-

established one (Lissitz, 2009), a desirable result to present on a scale development study was outside of the scope of this dissertation.

Pilot Study

The SCI-T pilot data collection occurred April 29 to July 15 2019. The instrument was available in Spanish and English on Qualtrics via separate links. The SCI-T Spanish was created to foster equity and representation in the access and response to the SCI-T, as well as increase study participation. TWIN-CS members received it via email two weeks after the initial email invitation to participate in the study. A total of 109 teachers and administrators accessed the SCI-T, of which 54 completed the full scale (i.e. provided responses to all 14 scenarios).

To provide further insight into the survey results, I collected qualitative data about the SCI-T and teachers' experience with sociocultural integration at their schools. Teachers participating at the TWIN-CS Summer Academy on June 23-26 who visited the Makers Space, were invited to draw scenarios from a fishbowl. They were asked to "think aloud" when completing the scenario –similar to the cognitive lab used in 2017. Four teachers completed the exercise, and thus feedback for only those four random scenarios were recorded.

Pilot Study Missing Data

The following descriptive statistics pertain to the combined responses of the SCI-T English and Spanish versions unless noted otherwise. One hundred and nine individuals accessed the survey, out of which 68 provided a response to at least one of the SCI-T's scenarios. Fifty-four participants (79%) provided responses to the full scale. Table 4.1 describes pilot study response levels.

Table 4.1*Pilot Study: Scale Completion Rates*

Response status	N	%
14 scenarios (full-scale)	54	79.4
13 scenarios	4	5.9
10-12 scenarios	2	2.9
9 or fewer scenarios	8	11.8
Total participants	68	100.0

Five out of sixty-eight participants who responded to at least one scenario accessed the Spanish language SCI-T. The remaining sixty-three (92.6%) accessed the English SCI-T. The average completion time for the sixty-eight participants who responded to at least one scenario was thirteen and a half minutes. This was a helpful value to use when advertising the survey to potential respondents in the main study. This average response time excludes six participants that had extremely high values (above 22 hours). Such values could occur when respondents access Qualtrics on more than one occasion to complete the survey.

Participants accessed the SCI-T via a link provided in an email, as well as a Quick-Response (QR) code given to attendees of the TWIN-CS 2019 Summer Academy. QR codes allowed attendees to view the survey on their mobile phones by scanning the picture with their mobile's camera. Out of the 109 responses recorded, eight (7.3%) accessed using a QR code. Across SCI-T languages, 20% of Spanish survey access was via QR code, while only 6.3% used this mode to access the English version. Out of the

eight surveys accessed via QR code, five participants answered one or more scenarios.

Table 4.2 summarizes the response status, by survey language, for participants who answered at least one of the 14 SCI-T scenarios.

Table 4.2

Pilot Study: Responses by Survey Language

Response Status	English		Spanish	
	n	%	n	%
Answered at least one scenario	63	100.0	5	100.0
<i>By Access Mode</i>				
Accessed via email	59	93.7	4	80.0
Accessed via QR code	4	6.3	1	20.0
<i>By Scale Completion</i>				
Completed full-scale	50	79.4	4	80.0
13 scenarios	4	6.3	0	-
10-12 scenarios	1	1.6	1	20.0
9 or fewer scenarios	8	12.7	0	-

Pilot Study Scenario Descriptive Statistics

Table 4.3 below presents the mean and standard deviation for each of the 14 scenarios for the sample of 68 teachers who responded to at least one scenario. Scenarios SCIT-1 and SCIT-2 (i.e., SID, PSE and APE, facets in High), which captures the high-level enactment of promotion of SCI, had the lowest means at 3.16 (SD = 0.62) and 3.04 (SD = 0.77) respectively. Conversely scenario SCIT-13 (i.e., SID, PSE and APE, facets in Low), which is a low-level scenario, had the highest mean at 4.3 (SD = 0.63). Comparing the averages to the response options and their corresponding values (i.e., “much less proactive” = 1, “about the same” = 3, “much more proactive” = 5) suggests

that the scenarios with high levels of enactment of SCI were harder for teachers to choose they were “much more proactive”. For scenarios with low levels of SCI enactment it means that teachers were more likely to choose that they were “much more proactive”. The descriptive statistics of all scenarios suggest that most are ordered according to their theoretical levels of SCI.

Table 4.3

Pilot Study: Scenario Descriptive Statistics

Scenario	M	SD	n	Order
SCIT-1_HHH	3.16	0.62	67	1
SCIT-2_HHH	3.04	0.77	67	3
SCIT-3_HMH	3.69	0.94	64	7
SCIT-4_HMH	3.35	1.06	63	8
SCIT-5_HHM	3.51	0.80	59	9
SCIT-6_HHM	3.46	0.68	59	13
SCIT-7_MMM	3.94	0.97	67	2
SCIT-8_MMM	4.12	0.67	59	14
SCIT-9_MLM	4.02	0.78	59	10
SCIT-10_MLM	3.92	1.10	64	6
SCIT-11_MML	4.07	0.77	58	12
SCIT-12_MML	3.84	0.93	67	4
SCIT-13_LLL	4.30	0.63	64	5
SCIT-14_LLL	4.05	0.96	58	11

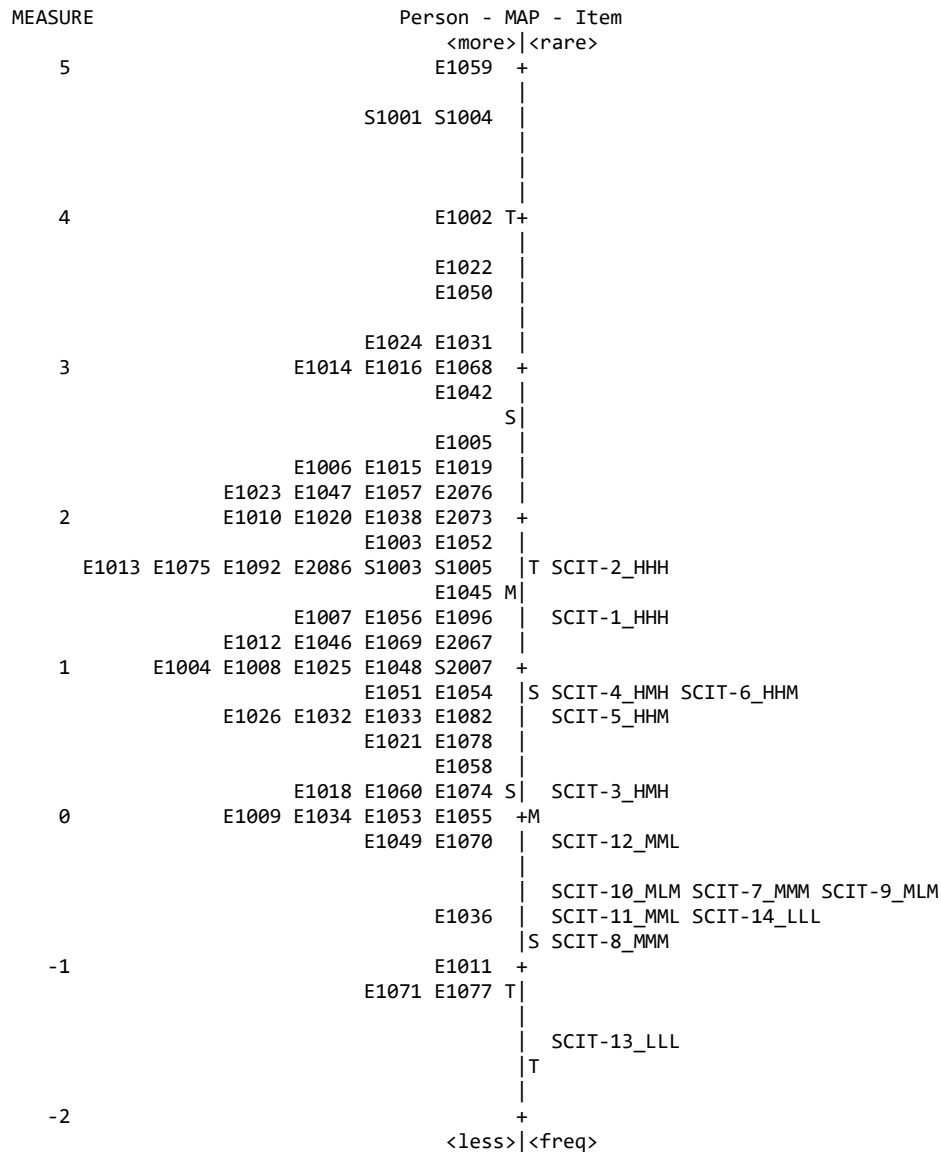
The number of responses by item shows a pattern in relation to the order of appearance. The last five scenarios in the scale have lower response rates than the scenarios shown first. Given that the sample is small in size for the number of scenarios, I included all cases in the Rasch analyses. The following subsections present the results of the pilot Rasch analyses for the full sample.

Pilot Study Variable Map (n=68)

The SCI-Teacher Scale's variable map in Figure 4.1 is the ladder-like empirical representation of the SCI continuum I described in theory. The variable map brings together scenario and teacher distributions on the same scale. Teachers are located on the left side of the vertical line. Teachers were assigned a five digit unique ID. The ID starts with the letter "S" or "E" depending on the language of the SCI-T they responded to; following a number "1" if they accessed the survey via email, or "2" if it was via QR code; followed by a three digit number capturing the order in which participants' responses were received. Scenarios are located on the right side of the vertical line on Figure 1. Each scenario is identified by a two-part name. First the survey name (SCI-T) followed by a number from one to fourteen, representing the order of appearance on the survey. Second, after the underscore, there are three letters representing the level of the scenario in each of the three facets ("L" is for low, "M" is for moderate, and "H" is for high) starting with self-identification, promotion of identity self-esteem, and agency for power equity. The variable map also displays three letters "M", "S", and "T". The letter M corresponds to the average score in logits separately for teachers and scenarios. The letter S shows one standard deviation distance from the teacher mean and the scenario mean, while the letter T shows two standard deviations. When looking at individuals, logits reflect how much teachers promote SCI at their school. For any given scenario, logits represent how difficult it is for teachers to say, "I am more proactive than teacher X". In this scale, a large positive number is a harder scenario, and the largest negative number is the easiest scenario.

Figure 4.1

Pilot Study: SCI-Teacher Scale Variable Map (n=68)



The variable map pointed to two main areas for improvement for the SCI-T. The first is related to the spread of the scenarios. Overall, scenarios tended to be lower on the continuum of SCI than respondents. This is confirmed by the graphical depiction of the mean for participants represented by the “M” on the left side of the variable map at

around 1.5 *logits*, while the mean of scenarios represented by the “M” on the right side of the variable map is located near zero. Scenarios SCIT-1 and SCIT-2 were designed to be very high representations of teacher promotion of SCI, but many participants considered themselves to be “much more proactive” than them. This first finding suggests that the highest representations of SCI promotion should be more committed and difficult to endorse. This observation is still valid considering that I expected TWIN-CS teachers to enact higher levels of SCI promotion than the average K-12 teacher. The distribution of SCI-T scenarios should go at least as high as the highest SCI promoting teachers. The second area for improvement is related to the redundancy in the level of estimated difficulty for some scenarios. There are two clusters of scenarios around -0.5 *logits* and 0.75 *logits*. In order to improve the spread and reduce the redundancy, some aspects were explored. First, I had to identify which items were theoretically expected to be at this level of difficulty (in terms of ranking since the actual difficulty estimate cannot be predicted *ex-ante*). The analysis of table 4.4 goes further into this point. Once I identified the scenarios that appeared in a different order than the theory suggested, I examined each facet’s representation in these scenarios. My hypothesis was that I would find facet representations that were perceived as much more difficult (or as much easier) to endorse than theorized. When targeting the redundancy and addressing the match between theoretical and estimated scenario difficulties, the spread and range were addressed too. The goal of the design, piloting, and revision process is to improve scenario difficulty distribution in each iteration, so that scenarios are spread out to map the SCI continuum; as well as having scenarios at each level that teachers are mapped onto.

Table 4.4 shows the difficulty and goodness of fit statistics obtained from the Rasch rating scale analysis using WINSTEPS (Linacre, 2020). The scenario difficulty estimates are in *logits*. Chapter 3’s review of the Rasch model provided an explanation of this unit. The mean difficulty was set to zero at estimation. Scenarios appeared in a random order consistently across the English and Spanish SCI-T surveys. The total number of responses by scenario ranges from 57 to 66. Our first statistic of interest is the measure column. *Measure* represents a scenario’s empirical difficulty level and a corresponding ordering in the variable map. We can see that roughly all scenarios have difficulties that match their hypothesized ordering, as illustrated by scenario names. For example, scenario SCIT-2 has high levels of facets SID, PSE, and APE (as signaled by the “HHH” in the scenario name). SCIT-2 is the most difficult scenario, or hardest for a teacher to respond they were “much more proactive” to. Similarly, SCIT-13 appears at the bottom of the table, which matches my expectation given it has low levels for each of three facets. Standard errors around these difficulty estimates have a mean of 0.20.

Table 4.4*Pilot Study: Difficulty Estimates and Fit Statistics*

Scenario	Measure	SE	INFIT MNSQ	OUTFIT MNSQ
SCIT-2_HHH	1.58	0.18	0.68	0.68
SCIT-1_HHH	1.30	0.18	1.03	0.98
SCIT-4_HMH	0.91	0.19	1.47	1.49
SCIT-6_HHM	0.76	0.20	0.62	0.63
SCIT-5_HHM	0.66	0.20	0.65	0.64
SCIT-3_HMH	0.13	0.19	0.95	0.93
SCIT-12_MML	-0.19	0.19	1.01	1.03
SCIT-10_MLM	-0.44	0.20	1.55	1.43
SCIT-7_MMM	-0.45	0.20	1.80	2.57
SCIT-9_MLM	-0.58	0.21	0.63	0.62
SCIT-14_LLL	-0.64	0.21	1.33	1.20
SCIT-11_MML	-0.72	0.22	0.75	0.72
SCIT-8_MMM	-0.87	0.22	0.87	1.17
SCIT-13_LLL	-1.47	0.22	0.72	0.95

Note. All statistics provided are in logits. MNSQ = mean square.

Examining the descending order of difficulty in Table 4.4, most scenarios match their theorized locations. For example, scenarios SCIT-2 and SCIT-1 are the hardest to say “I am more proactive than Teacher X” to. This is an expected result as shown by the three “H”s in the scenario names that show that all three facets are at their highest level of sociocultural integration. Scenarios with two facets at level high and one at moderate were the next in terms of difficulty. Scenarios with facets at the moderate level appear next, and at the bottom is the easiest scenario to endorse, SCIT-13, with all facets at the lowest level. Two scenarios appear out of their expected ordering. The first is SCIT-7, which is easier to say “I am more proactive than” than the theoretically easier SCIT-12

and SCIT-10. The second is scenario SCIT-14, which appears more difficult to say “I am more proactive than”, compared to other low SCI scenarios (e.g. SCIT-13).

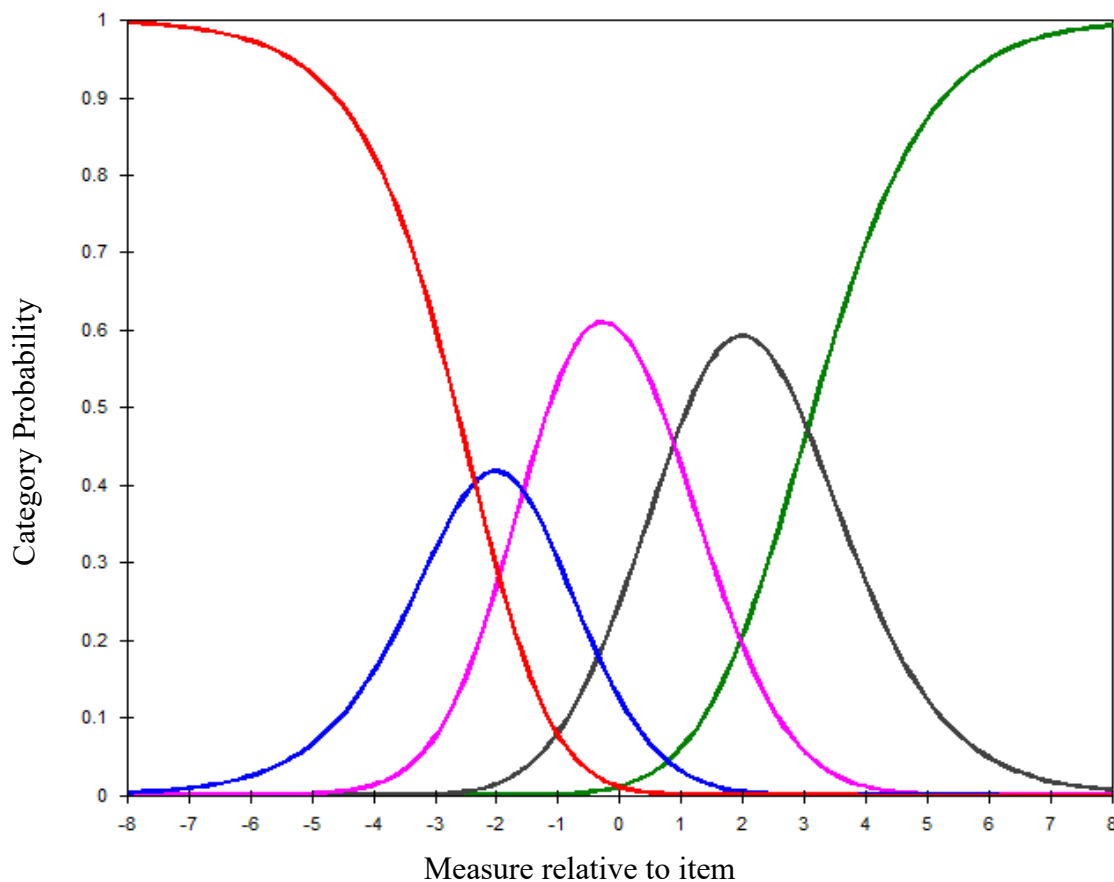
The OUTFIT and INFIT columns are goodness-of-fit statistics for scenarios and teachers produced by WINSTEPS. INFIT is a weighted mean-square statistic, and OUTFIT is an unweighted mean square statistic, both are based on the differences between teachers observed responses to the SCI-T and their expected responses under the Rasch rating scale model (Ludlow, 1983; Wright & Masters, 1982). These statistics inform how appropriately teachers are responding to scenarios. OUTFIT is sensitive to unusual responses to as few as a single scenario by any given teacher in relation to their estimated SCI level, helping identify outliers. INFIT, on the other hand, is sensitive to uncharacteristic response patterns by teachers to scenarios located around their estimated level of SCI. Table 7 reports the mean square INFIT and OUTFIT statistics. The thresholds for acceptable unstandardized INFIT and OUTFIT mean squares differ by authors. Enterline et al. (2008) argue that the majority of studies use a range of 0.7 to 1.3. Ludlow et al. (2008; 2014) suggest values between 0.5 and 1.5, and a more liberal 1.2 threshold “to avoid missing potential problems in the initial stage” (Chang, 2017, p. 110). Using Ludlow et al.’s (2008; 2014) thresholds, scenarios SCIT-7, (INFIT 1.80), SCIT-10 (INFIT 1.55) and SCIT-4, (INFIT 1.47) show some misfit. Out of these scenarios, only SCIT-7 appears to be at a lower level of difficulty than expected given its facets’ levels. I flagged these scenarios for revision to prevent potential issues in the main study.

The category characteristic curves show the likelihood of a teacher responding in each response category based on the difference between a given teacher’s SCI support estimate and a given scenario’s difficulty (Wright & Masters, 1982). Figure 4.2 shows the

SCI-T scale's category curves. As a scale, the SCI-T five-category rating scale appears adequate in terms of separation based on the distance between successive category peaks at different points of the spectrum. The probability to answer "much less proactive than teacher X", through "much more proactive than teacher X" shows a peak at different levels of teacher's support of SCI. This means that the category threshold estimates show the desirable property of increasing from the first category to the fifth (-2.36, -1.57, 0.88, 3.05). In terms of dispersion, category b) "I would be less proactive" (the blue curve), has the lowest peak and corresponding area where it is has the highest probability of response. A possible course of action was to revise the language of option b) before the main study. However, that would require option d) "I would be more proactive" to be revised too to preserve their perceived equivalency in the five-point rating scale. After consulting with my dissertation committee, I decided against this revision.

Figure 4.2

Pilot Study: Scenario Response Category Characteristic Curves



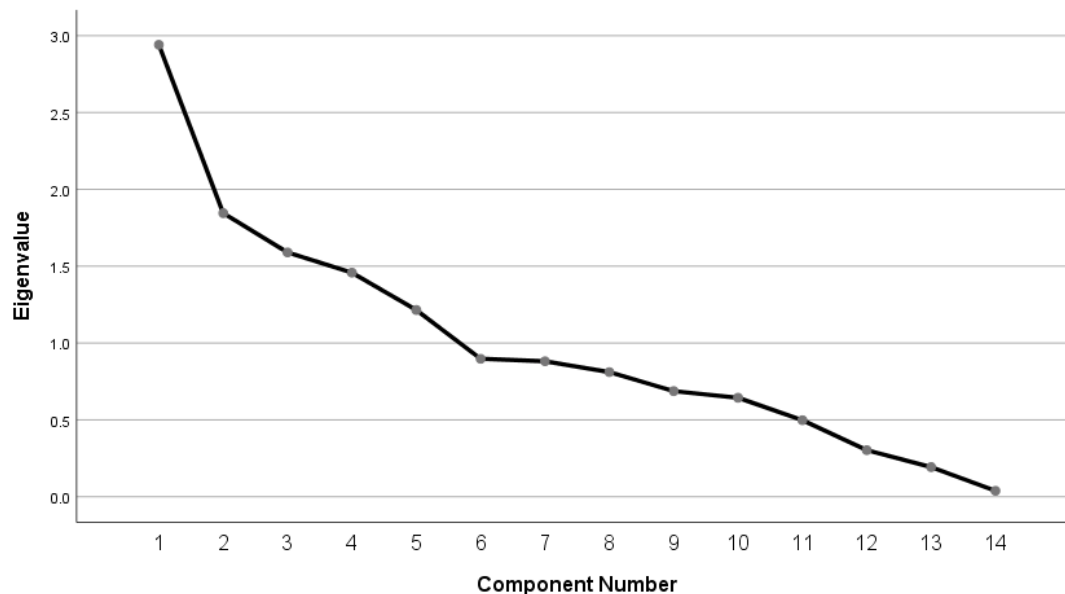
The analysis of teacher responses uses the same criteria for flagging potentially troublesome teacher response patterns as do scenarios. However, Andrich & Marais (2019) ask researchers to use fit statistics in context and not absolutely. To follow this advice, I include a table in Appendix 8 where teachers are ordered by their INFIT MNSQ, and then their OUTFIT MNSQ, to assess more clearly how the values change. If there are big differences or a large jump in the values from one teacher to the next, these could suggest those profiles are of concern. While teacher E1059 showed an atypical

response pattern and high INFIT and OUTFIT, neither their survey taking time nor their responses seem unreasonable. In addition, it is expected under normal probability theory to observe 5% of responses that are outside of two standard deviations from the mean of the sample. As a result, the person's data record remained in the analysis.

Finally, I evaluated the assumption of unidimensionality for the SCI-T scale. First, I analyzed the rating scale model's residual matrix using SPSS. The residual matrix's rows are the teachers who answered the SCI-T, and the columns contain the difference between their observed score on the SCI-T versus their expected score. The estimated Rasch model assumes that after removing the contribution of sociocultural integration from responses, all that is left is random noise. As a result, conducting PCA on residuals from a unidimensional data set is expected to extract no principal components (Wright, 1996). Figure 4.3 presents one standard PCA check for dimensionality. The scree plot is used as a visual "test". Eigenvalues are plotted on the y-axis against the number of principal components on the x-axis. The "test" is to observe when the negative trend in the slope starts to become flat, and the number of components for the analysis is given where there is a discernible elbow. For random data that exhibit zero valued inter-correlations, the slope of the scree plot would essentially be flat at approximately an eigenvalue of one for all components. Figure 4.3 shows there is a large drop in the eigenvalues after one principal component. This is evidence to suggest that there is a lack of unidimensionality in the raw data (Cattell, 1966). That is, unexpected correlated variance remains in the residuals.

Figure 4.3

Pilot Study: Residual Matrix Principal Component Analysis Scree Plot



In order to determine whether a principal component analysis (PCA) of the residual matrix was warranted, I looked at critical statistics like the determinant. If the value of determinant is too low ($<.00001$) there may be multicollinearity or singularity (i.e. the items are too related); but a non-zero value signals that there is variability to warrant PCA. For the SCI-T residuals the determinant was 0.005, a sign in favor of further analysis. Other statistics explored were the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy, and Bartlett's Test of Sphericity. The KMO value was 0.132, which is a sign that the pattern of correlations isn't particularly strong so that a PCA is perhaps not necessary. The p-value for Bartlett's Test was below 0.05, suggesting that there are relationships in the residual matrix that should be explored. Based on the scree plot and Bartlett's Test results, I conducted a principal component analysis (PCA) on the residual matrix.

Table 4.5 presents variance explained by each of the extracted components by PCA. The first component accounts for approximately 21% of the residual variance and has an eigenvalue of 2.94, which is above the cutoff of one used by the Kaiser-Guttman rule to determine the number of latent traits (Guttman, 1954; Kaiser, 1960). While there is more than one component with an eigenvalue above one, the variance explained for each is below 13%, below the 20% cutoff to be treated as “significant” (Zopluoglu & Davenport Jr., 2017). A more careful look at the first component is required since the shared variance is high. The scenarios that are contributing more heavily to shared variance of this component, ordered from high to low according to their loadings in parenthesis, are SCIT-2 (-0.7), SCIT-1 (-0.7), SCIT-10 (0.6), SCIT-11 (0.6), SCIT-14 (0.6), and SCIT-9 (0.5). The remaining scenarios had loadings below ± 0.4 , so I disregarded them. There is a mix of very high and very low difficulty in this list of scenarios. One could think about why the residuals of these extreme items stand out. Surprising responses, which is how I characterize these observations, are responses for which the residuals are high in the positive or negative direction (i.e. observed response is largely different from the expected). This could only occur with scenarios at the extremes. For SCIT-1 and SCIT-2, there are large positive differences, meaning that some individuals with low SCI were selecting “much more proactive” and “more proactive”. The reverse is true for the moderate/low scenarios in the list, some teachers with high SCI were selecting “much less proactive” and “less proactive” when their SCI estimates corresponded to higher categories. In this way, the large residuals in extreme scenarios is not necessarily an indication of latent construct. It’s more of a byproduct of the scale design. Included in Appendix 9 is the component plot of the residual variance in

rotated space, which shows how closely related scenarios are to one another and to the two principal components extracted. The result is a random-looking scatter of the residuals, supporting the unidimensionality assumption in the raw data. Note, however, that in other RGS methodology studies (Chang, 2017), the residuals of extreme scenarios were highly correlated – as we see here for SCIT-1 and SCIT-2.

Table 4.5*Pilot Study: PCA of Residuals Analysis – Total Variance Explained*

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.94	21.01	21.01	2.94	21.01	21.01
2	1.85	13.18	34.18	1.85	13.18	34.18
3	1.59	11.35	45.54	1.59	11.35	45.54
4	1.46	10.41	55.95	1.46	10.41	55.95
5	1.22	8.68	64.63	1.22	8.68	64.63
6	0.90	6.41	71.04			
7	0.88	6.30	77.33			
8	0.81	5.79	83.13			
9	0.69	4.91	88.03			
10	0.64	4.60	92.63			
11	0.50	3.56	96.19			
12	0.30	2.16	98.35			
13	0.19	1.38	99.73			
14	0.04	0.27	100.00			

SCI-T Revisions informed by Pilot Study Conclusions

The pilot study shows that twelve out of the fourteen scenarios in the SCI-T were empirically ordered consistently with the theory presented in this dissertation. I detected two areas of improvement after looking at the results. In total, I had three areas of improvement taking into account feedback received at my proposal hearing. The first weakness was related to the spread of scenarios along the SCI spectrum. As seen in the variable map, the most difficult scenario (SCIT-2) is much lower than teacher E1059, who has the highest SCI. In fact, there are no scenarios accompanying the top third of the highest SCI teachers. To a certain extent, this was an expected result. The teachers who attended the TWIN-CS Summer Academy work in schools where the promotion of sociocultural integration is an explicit goal, compared to an average non-bilingual secular school. However, the SCI-T has to measure these very high SCI teachers well. Therefore, to remedy the absence of scenarios at the top of the distribution, I revised the language of the most difficult and moderate scenarios to make them even harder to endorse. As a result, I should see scenarios at the top of the variable map in the main study, as well as more spread in the middle of the SCI-T spectrum as represented in the variable map. Secondly, I identified two items that appear out of the theorized order. SCIT-7, which is easier to say “I am more proactive than” than the theoretically easier SCIT-12 and SCIT-10; and SCIT-14, which appears more difficult to say “I am more proactive than”, compared to other low SCI scenarios (e.g. SCIT-13). Modifications in the language that could align them empirically with their theoretical difficulty were made for the main study. The third and final scenario improvement needed was to include more facets on the racial dimension in order to match the target population of the SCI-T (rather than

make the population of interest narrower). The SCI-T tested in the pilot study included fourteen scenarios that were more skewed towards cultural, ethnic and linguistic diversity.

Main Study

The data collection for the main study took place between December 15, 2020 and February 15, 2021. Teachers were recruited using *snowball sampling*, which is an outreach strategy to locate participants with certain characteristics who are difficult to find (Sadler et al., 2010). This was particularly true for the K-12 teacher population during the COVID-19 pandemic, whose work became remote or partially remote and for whom exhaustion was a real barrier to participating in research. While it is a non-probabilistic method that is likely to result in a non-representative sample, the circumstances of strain in the population of interest was an appropriate strategy to fulfill the objective of this data collection to provide evidence of the SCI-T psychometric properties. Following the snowball technique, I contacted a few stakeholders and teachers, and asked them to invite other teachers to participate in my study. My initial contacts used their social and/or professional networks to mimic “a process analogous to a snowball rolling down a hill” (Wasserman et. al, 2005 as cited in Sadler et al., 2010) in a “semi-self-directed, chain-referral, recruiting mechanism able to reach the hard-to-reach target group in a more pragmatic and culturally competent way.” (Sadler et al., 2010, p. 3).

Main Study Missing Data

One hundred and five participants responded to all 14 scenarios of the SCI-T. Seventy percent of respondents completed 13 scenarios or the full scale, suggesting the

recruitment strategy was successful. The total of responses available for the analysis complies with the recommendation by DeVellis (2003) of a minimum of 10 responses per scenario, as seen on Table 4.6.

Table 4.6

Main Study: Scale Completion Rates

Response Status	n	%
14 scenarios (full-scale)	105	64.4
13 scenarios	9	5.5
10-12 scenarios	6	3.7
9 or fewer scenarios	43	26.4
Total participants	163	100.0

Main Study Sample and Scenario Descriptive Statistics

Table 4.7 provides an overview of the teacher respondents' characteristics. Participants were allowed to skip questions and the demographics section altogether, resulting in responses ranging from 101 to 115 for different characteristics. In addition, the summary of characteristics is presented for the full sample (i.e. all teachers who provided the information), and for the group who completed the full scale (i.e. provided responses to the 14 scenarios). For teachers who provided their age range, 45.2% were in the 29 years old or younger. The sample skewing young is not surprising given that many of the seed contacts in the recruitment were young teachers or associations of young graduates/teacher candidates. The following age group, 30-39, had the largest number of respondents for a single category at 41 participants (35.7%). The remaining 6.9% of participants who provided age were 50 years or older. While the recruitment strategy was not intended to

yield a representative sample, the age distribution of participants mirrors that of nationwide public school teachers, with the exception of teachers with over 20 years of experience (corresponding to the two last age groups) which accounted for 22% in 2015-2016 (NCES, 2019). In terms of grade level taught, 42.5% of teachers declared K-5, and roughly 47% teach in K-8. Almost 18% of participants teach middle school, and 30% teach high school. Twelve teachers selected more than one category, representing 10.6% of the sample. Regarding participants' race/ethnicity, 81% identify as White, 10.4% as Hispanic, 4% as African American, 3.5% as Asian, and 0.9% as mixed race. The breakdown on this characteristic closely mirrors NCES' (2019) distribution for the 2015/16 school year.

Table 4.7*Main Study: Teacher Characteristics*

Characteristics	Full sample		Completed 14 scenarios	
	n	%	n	%
<i>Age</i>	<i>115</i>	<i>100.0</i>	<i>102</i>	<i>100.0</i>
Under 25	15	13.0	13	12.7
25-29	37	32.2	33	32.4
30-39	41	35.7	36	35.3
40-49	14	12.2	13	12.7
50-59	3	2.6	2	2.0
60 or older	5	4.3	5	4.9
<i>Grade level taught</i>	<i>113</i>	<i>100.0</i>	<i>101</i>	<i>100.0</i>
K-5	48	42.5	41	40.6
K-8	5	4.4	4	4.0
K-12	2	1.8	2	2.0
6-8	20	17.7	19	18.8
6-12	5	4.4	5	5.0
9-12	33	29.2	30	29.7
<i>Race</i>	<i>115</i>	<i>100.0</i>	<i>102</i>	<i>100.0</i>
Hispanic	12	10.4	10	9.8
White Non-Hispanic	93	80.9	82	80.4
Asian	4	3.5	4	3.9
Black or African American	5	4.3	5	4.9
American Indian or Alaska Native	0	-	0	-
Native Hawaiian or other Pacific Islander	0	-	0	-
More than one racial category chosen	1	0.9	1	1.0

Table 4.8 presents the mean and standard deviation for each of the 14 scenarios corresponding to the final version of the SCI-T. Scenarios are ordered in descending order of theoretical SCI promotion. The higher the scenario is on the table, the more difficult it is (i.e. teachers should be less likely to say that they are “more proactive” than the teachers described). Overall, the scenario averages correspond with their theoretical levels. Similar to the pilot study findings, scenarios SCIT-1 and SCIT-2 (i.e., SID, PSE

and APE, facets in High), which capture the high-level enactment of promotion of SCI, had the lowest average at 2.78 (SD = 0.75) and 2.54 (SD = 0.81) respectively. The averages for both scenarios are lower than in the pilot study, suggesting that the changes made to the scale to make the highest enactments of SCI to be more difficult for a teacher to endorse were successful. On the lower end of the SCI continuum, scenario SCIT-13 (i.e., SID, PSE and APE, facets in Low) had the highest mean at 4.62 (SD = 0.91), followed by scenarios SCIT-11 (M=4.34, SD=0.87) and SCIT-14 (M=4.32, SD=0.85). An average value above four means that teachers were more likely to say that they were “more proactive”, or “much more proactive”, than the hypothetical teacher described in the scenario. Consistent with scale completion rates, the range of the responses across scenarios varies from 115 to 162.

Table 4.8

Main Study: Scenario Descriptive Statistics

Scenario	M	SD	n
SCIT-1_HHH	2.78	0.75	162
SCIT-2_HHH	2.54	0.81	150
SCIT-3_HMH	3.11	0.55	132
SCIT-4_HMH	2.86	1.03	124
SCIT-5_HHM	3.52	0.72	122
SCIT-6_HHM	3.18	0.67	115
SCIT-7_MMM	3.70	0.88	150
SCIT-8_MMM	3.89	0.79	118
SCIT-9_MLM	4.15	0.80	120
SCIT-10_MLM	3.95	0.83	131
SCIT-11_MML	4.34	0.87	117
SCIT-12_MML	3.83	1.29	141
SCIT-13_LLL	4.62	0.91	137
SCIT-14_LLL	4.32	0.85	117

Main Study Variable Map

Figure 4.1 represents the distribution of scenario difficulties for the pilot study. The revisions required for the SCI-T were evident: there was scenario redundancy (many scenarios at the same difficulty level), and the spread of scenarios from the highest to the lowest was truncated in comparison to the distribution of teachers' SCI. The variable map for the main study is presented in Figure 4.4. Based on the data for 164 participants, the scenarios are located along the continuum of sociocultural integration according to theory. Scenarios with higher levels of the three facets of SCI are positioned higher along the central line of the figure, than those with lower levels. While the highest scenario (SCIT-2) is harder to endorse (i.e. for a teacher to say they are "more proactive than"), there are still teachers at higher levels of SCI than this scenario. While the revisions worked for a distribution of SCI enactments that is more spread and similar to respondents', the more difficult representations of SCI as described in scenarios are still easier than teachers consider they would enact.

The scenarios that are theoretically at the same levels, appear in close proximity confirming that different enactments are indeed at the high, moderate, or low levels of SCI promotion –and are perceived accordingly regardless of the combinations (e.g. a MML and MLM scenario should appear close).

Main Study Rating Scale Model Results

Table 4.9 shows the difficulty and goodness of fit statistics obtained from the Rasch rating scale analysis using WINSTEPS. As in the pilot, scenario difficulty estimates are in *logits* and the mean difficulty was set to zero at estimation. Examining the descending order of difficulty in Table 4.9, most scenarios match their theorized locations. For example, scenarios SCIT-2 and SCIT-1 are the hardest to say “I am more proactive” to. This is an expected result as shown by the three “H”s in the scenario names that show that all three facets are at their highest level of sociocultural integration. Scenarios with two facets at the high level and one at moderate were the next more difficult to endorse. For this group it is interesting to note that scenarios with “high” levels of facets self-identification (SID) and agency for power equity (APE) were more difficult than those with “high” levels of SID and promotion of self-identification (PSE). Scenarios with moderate level facets appear next, and at the bottom is the easiest scenario to endorse, SCIT-13, with all facets at the lowest level. Two scenarios appear out of their expected ordering. The first is SCIT-8, which is easier to say “I am more proactive than” than the theoretically easier SCIT-12. The second is scenario SCIT-11, which appears easier to say “I am more proactive than”, compared to the theoretically easier SCIT-14.

Table 4.9*Main Study: Difficulty Estimates and Fit Statistics*

Scenario	Measure	SE	INFIT		OUTFIT	
			MNSQ	ZSTD	MNSQ	ZSTD
SCIT-2_HHH	1.96	0.11	1.13	1.11	1.35	2.85
SCIT-1_HHH	1.52	0.11	0.95	-0.40	1.00	0.05
SCIT-4_HMH	1.45	0.12	1.41	2.9	1.41	2.90
SCIT-3_HMH	0.99	0.12	0.77	-1.97	0.80	-1.72
SCIT-6_HHM	0.90	0.12	0.58	-3.77	0.58	-3.82
SCIT-5_HHM	0.30	0.12	0.77	-1.98	0.78	-1.94
SCIT-7_MMM	-0.11	0.11	0.76	-2.36	0.78	-2.19
SCIT-12_MML	-0.34	0.12	1.97	7.00	1.91	6.56
SCIT-8_MMM	-0.39	0.13	0.60	-3.84	0.64	-3.37
SCIT-10_MLM	-0.54	0.12	0.77	-2.20	0.08	-2.32
SCIT-9_MLM	-0.89	0.13	0.62	-3.58	0.64	-3.20
SCIT-14_LLL	-1.29	0.14	0.81	-1.52	0.73	-2.03
SCIT-11_MML	-1.34	0.14	1.02	0.19	1.01	0.15
SCIT-13_LLL	-2.21	0.16	2.09	5.75	1.42	2.08

Table 4.9 also contains the weighted (INFIT MNSQ) and unweighted (OUTFIT MNSQ) mean squares of standardized residuals between the expected and observed responses. A desirable outcome is a value between ± 2 for the standardized residuals' t-statistic (ZSTD) corresponding to each misfit indicator. Following the pilot study strategy, I use the "forgiving bounds" 0.5 and 1.5 to identify scenario misfit. Scenarios SCIT-12 (INFIT MNSQ 1.97, OUTFIT MNSQ 1.91) and SCIT-13 (INFIT MNSQ 2.09) show misfit according to the criteria. These scenarios also had t-statistics (ZSTD) for INFIT and OUTFIT larger than 2.

Scenario misfit can be related to some extent to extreme responses. In order to identify any potential outliers, I present the results of the respondent fit statistics in Table 4.10. Only respondents with INFIT or OUTFIT MNSQ values above 2 are included in the

table, amounting to about 10% of total respondents. It is expected under normal probability theory to observe 5% of responses that are outside of two standard deviations from the mean of the sample. However, there is a large number of respondents flagged with misfit, and with very large INFIT MNSQ values.

Table 4.10

Main Study: Selected Respondent Fit Statistics (n=14)

Respondent ID	Measure	INFIT		OUTFIT	
		MNSQ	ZSTD	MNSQ	ZSTD
R_3PY	-0.11	8.44	8.74	8.48	8.76
R_1Qn	-1.8	8.23	4.93	9.90	5.88
R_UoO	-0.5	4.97	5.94	5.04	6.02
R_3k0	0.5	4.68	5.24	5.25	5.75
R_UMS	-1.13	4.15	2.24	4.25	2.31
R_29g	0.4	3.96	4.99	4.35	5.40
R_3KA	-1.6	3.66	4.48	4.20	5.03
R_3sy	2.06	3.14	3.78	3.71	4.14
R_An5	-0.6	2.48	1.75	2.38	1.70
R_2sR	-0.65	2.22	2.43	2.16	2.36
R_3ES	-0.79	2.05	1.88	2.21	2.09
R_1eD	0.02	2.04	2.35	2.05	2.36
R_2yd	0.82	2.02	1.79	2.31	2.14
R_1kL	1.32	2.01	2.28	1.86	1.97

I decided to consider for deletion the seven participants with weighted and unweighted mean square residuals (i.e., INFIT- and OUTFIT-MNSQ > 3.5) and associated T-statistics (ZSTD > 2), using less forgiving standards than suggested by Linacre (2016) and Smith et al. (2003). Participants R_1Qn and R_UMS completed five and two scenarios (out of fourteen) respectively. Their misfit is likely due to the limited

information available to estimate SCI levels reliably. The five remaining participants completed the fourteen scenarios, and three of them provided demographic information. Two of the participants identified as Non-White, and one as Non-Hispanic White. Participant R_3PY showed misfit in 9 out of 14 scenarios, in half of them selecting response categories above expectations and the other half below. This response pattern seems like the participant was perhaps unengaged and selected responses at random. Participants R_3KA and R_UoO show misfit in 5 out of 14 scenarios. R_3KA responded to the survey in three minutes making their pattern suspicious for randomness; while R_UoO took an average amount of time. Participant R_29g is a moderate SCI level teacher whose observed responses for scenarios SCIT-7, SCIT-12, and SCIT-13 were below the expected (standardized residuals of -3, -3, and -5 respectively). Finally, participant R_3k0 was at a moderate level of SCI whose observed responses for scenarios SCIT-1, SCIT-4, and SCIT-5 were above the expected (standardized residuals of -3, -3, and -5 respectively). It was interesting to find that participant misfit occurred for six out of seven participants in item SCIT-13 (all with negative standardized residuals), and for five out of seven in item SCIT-1 (all with positive standardized residuals). The sign of the residuals make sense because SCIT-13 is a low facet combination scenario, and SCIT-1 is the exact opposite. I decided to delete these participants and reproduce analyses with the remaining 156 participants. Eliminating respondents is not capricious. "Careless responses", how Meade and Craig refer to random and nonrandom inattentive responses (2012), add to the error variance, in turn attenuating correlations, reducing internal consistency reliability estimates, and potentially contaminating the psychometric results (Meade & Craig, 2012). It is not surprising to come across careless responses in a

Qualtrics based survey data collection, as authors suggest that the inability to control the survey taking environment makes this a more pervasive issue (Buchanan, 2000).

Table 4.11 summarizes some important model statistics pre and post deletion of observations. Item separation gives an indication of the item hierarchy, the larger the value signals that the scenario difficulty hierarchy supports the validity of sociocultural integration as a construct. Both person and item separation are larger after the deletion of careless responses. Person reliability is also larger, meaning that it is much more likely that teachers' SCI levels are correctly assigned (i.e., teachers with high SCI are being estimated as having higher measures). The person and item raw score-to-measure correlations are the values of the Pearson correlation between response values and the difficulty estimates, values close to 1.0 are more desirable. In this case, the post-deletion model has lower person but higher item correlation. Model reliability estimates, Cronbach Alpha and 50-item reliability, are slightly larger or unchanged. Overall, I see slightly better fit statistics after deletion.

Table 4.11

Main Study: Model Fit Statistics Before and After Deletion

Model Statistics	n=163	n=156
Item separation	8.46	10.46
Item raw score-to-measure correlation	-0.84	-0.87
Person separation	1.23	1.40
Person reliability	0.60	0.66
Person raw score-to-measure correlation	0.43	0.37
Cronbach Alpha (K-20) person raw score reliability	1.00	1.00
Standardized (50 item) reliability	0.89	0.90

Figure 4.5 shows the variable map for the sample after deletion in order to assess any changes in the ordering of scenarios and to visualize the separation in the difficulty estimates. Compared to Figure 4.4, scenarios are more separate (e.g., SCIT-1 and SCIT-4 are not at the same difficulty level), and while the top of distribution of scenario difficulties is lower than the teachers', SCIT-2 is now at a higher level.

Table 4.12 shows the difficulty and goodness of fit statistics for scenarios after deletion. Confirming the changes suggested by Figure 6, scenario difficulties have changed slightly as a result of the deletion of respondents showing misfit. SCIT-2 up to SCIT-8 have higher difficulties than in the previous estimation. SCIT-12 is now located after SCIT-8, in correspondence with the theoretical level of both scenarios. SCIT-11 has not changed location, and is still appearing as an easier to say “much more proactive” than in theory. SCIT-13 is consistently the easiest scenario, but it is worth noting that it now has a lower difficulty estimate. In terms of misfit, using the 0.5-1.5 thresholds, I flagged scenarios SCIT-12 (INFIT MNSQ 2.23, OUTFIT MNSQ 2.16) and SCIT-4 (INFIT MNSQ 1.57, OUTFIT MNSQ 1.57). These scenarios also had t-statistics (ZSTD) for INFIT and OUTFIT larger than 2. After the deletion, SCIT-13 is no longer showing misfit.

Table 4.12*Main Study: Difficulty Estimates and Fit Statistics After Deletion*

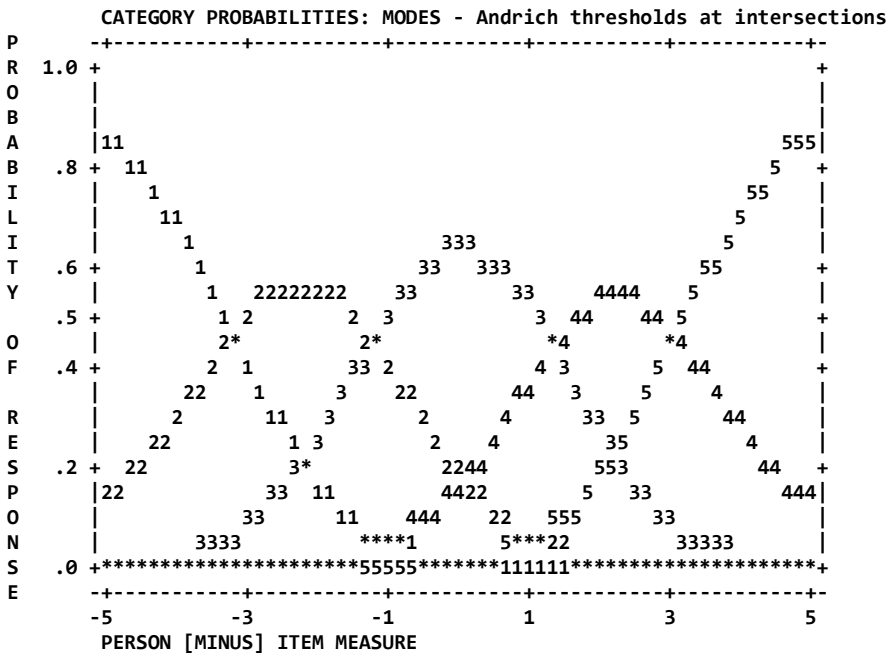
Scenario	Measure	SE	INFIT		OUTFIT		n
			MNSQ	ZSTD	MNSQ	ZSTD	
SCIT-2_HHH	2.66	0.13	1.02	0.2	1.04	0.38	144
SCIT-1_HHH	2.10	0.12	0.88	-1.05	0.88	-1.06	155
SCIT-4_HMH	1.96	0.14	1.57	3.73	1.57	3.71	119
SCIT-3_HMH	1.31	0.13	0.83	-1.38	0.83	-1.38	127
SCIT-6_HHM	1.21	0.14	0.54	-4.00	0.54	-4.07	110
SCIT-5_HHM	0.44	0.14	0.82	-1.45	0.82	-1.44	117
SCIT-7_MMM	-0.20	0.13	0.82	-1.71	0.84	-1.47	143
SCIT-8_MMM	-0.49	0.14	0.7	-2.74	0.75	-2.19	113
SCIT-12_MML	-0.57	0.13	2.23	8.23	2.16	7.71	135
SCIT-10_MLM	-0.70	0.14	0.90	-0.83	0.89	-0.87	126
SCIT-9_MLM	-1.09	0.15	0.66	-3.09	0.70	-2.51	116
SCIT-14_LLL	-1.64	0.16	0.71	-2.44	0.67	-2.41	112
SCIT-11_MML	-1.72	0.16	0.88	-0.89	1.00	0.07	113
SCIT-13_LLL	-3.27	0.22	1.13	0.74	1.02	0.17	131

Figure 4.6 provides a visual representation of the distribution of response categories. Numbers one through five represent each response category, where “1” is “much less proactive”, “2” is “less proactive”, “3” is “about the same”, “4” stands for “more proactive”, and “5” is “much more proactive”. Ideally, each response category has a range in the X-axis, which represents the person minus scenario difficulty, where it is dominant (i.e., it should look like a series of hills with peaks at each response category). This expectation is confirmed in Figure 4.6. Each number has a space along the SCI promotion level (X-axis) where it is the most probable response category. As such, for lower levels of SCI promotion, category “1” has the highest probability, and “5” is at the highest probability in the opposite end. This is evidence that all response categories were

used, and that these align with their theoretical position. For example, for higher SCI promotion, I expect to see the probability of choosing “much more proactive” to any scenario go up.

Figure 4.6

Main Study: Cumulative Probability Map by Andrich Thresholds



Next, I evaluated the assumption of unidimensionality for the main study SCI-T scale using two strategies. First, I conducted a factor analysis (FA) to confirm that the variability in the data is due to one latent variable: sociocultural integration. FA of response data is a dimension reduction technique that extracts from a given number of observed variables (in this case 14 scenarios) a smaller number of factors (Worthington & Whittaker, 2006). Exploratory FA is used to provide evidence of construct validity when developing a scale. In an exploratory FA of the SCI-T responses, I was expecting to find an underlying unidimensional structure with one factor explaining a considerable amount of variation in the data. The second strategy was to conduct a PCA analysis of the residuals matrix, similar to the pilot study.

The results from the exploratory FA show that there is a meaningful underlying factor. The determinant was 0.01; the KMO was 0.82 very close to 1, which is the desirable value when looking for an underlying latent variable; and Bartlett's Test was significant. Table 4.13 summarizes FA results. One component was extracted with an eigenvalue of 4.7, explaining a total of 33.77% of the variance in scenario responses. I take this result to suggest that the underlying latent trait being captured is SCI. I disregarded the remaining components as they have eigenvalues lower than 2, and explained variance around 10%.

Table 4.13*Main Study: PCA of Responses – Variance Explained*

Factors	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.73	33.77	33.77	4.73	33.77	33.77	3.84	27.45	27.45
2	1.78	12.71	46.48	1.78	12.71	46.48	2.21	15.76	43.20
3	1.20	8.59	55.07	1.20	8.59	55.07	1.66	11.87	55.07
4	0.91	6.48	61.56						
5	0.85	6.07	67.63						
6	0.73	5.21	72.83						
7	0.69	4.95	77.78						
8	0.62	4.46	82.24						
9	0.57	4.08	86.32						
10	0.47	3.38	89.70						
11	0.44	3.15	92.85						
12	0.41	2.92	95.76						
13	0.34	2.41	98.17						
14	0.26	1.83	100.00						

The second strategy uses the same procedure but a different data source: the matrix of residuals. The objective of the PCA of residuals is to test whether, after removing the shared variance due to teachers' sociocultural integration level from their responses, all that is left is random noise. The ideal result in this case is to find that there are no components with large eigenvalues and explaining a moderate percentage of the variance. In order to determine whether a PCA of the residual matrix was warranted, I checked the values of the determinant, KMO, and Bartlett's Test. In order to have a basis of how the results improved from the pilot to the main study, I compare selected statistics. The determinant was 0.007, lower than in the pilot study but still suggesting PCA was appropriate. The KMO was 0.04, lower than the pilot study's; but the p-value for Bartlett's Test was below the 0.05 threshold. All three statistics suggest that further analysis is required. Table 4.14 presents variance explained by each of the extracted components by PCA. The first component accounts for approximately 18% of the variance and has an eigenvalue of 2.6, down from the 21% of variance explained in the pilot study residuals – while lower it is still considerable and warrants further examination. There is only one component with an eigenvalue above two, so I will disregard the others.

Table 4.14*Main Study: PCA of Residuals Analysis – Total Variance Explained*

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.55	18.23	18.23	2.55	18.23	18.23
2	1.63	11.60	29.83	1.63	11.60	29.83
3	1.29	9.20	39.03	1.29	9.20	39.03
4	1.28	9.17	48.21	1.28	9.17	48.21
5	1.21	8.62	56.82	1.21	8.62	56.82
6	1.19	8.53	65.35	1.19	8.53	65.35
7	0.96	6.89	72.24			
8	0.90	6.44	78.68			
9	0.77	5.47	84.15			
10	0.67	4.79	88.94			
11	0.62	4.45	93.39			
12	0.52	3.71	97.10			
13	0.39	2.82	99.92			
14	0.01	0.09	100.00			

A more careful look at the first component, summarized in Table 4.15, reveals that there are three scenarios with loadings above ± 0.4 . Scenarios SCIT-1 (0.8) and SCIT-3 had a large positive relationship to the component, while SCIT-12 (-0.7) had a large negative relationship to the component. Unlike the pilot study's residual analysis, there is not an easily discernible pattern for the three scenarios that contribute to the components, suggesting that it may be random noise. Similar to the pilot, one high level SCI scenario (SCIT-3) has a large positive relationship, meaning that teachers are choosing more positive response categories; and those response patterns are correlated with teachers choosing less positive response categories than expected to lower and more moderate scenarios (SCIT-8 and SCIT-12). Included in Appendix 10 is the component plot in rotated space, which shows how closely related scenarios are to one

another and to the two largest principal components extracted. The result is a random-looking scatter of the residuals, supporting the unidimensionality assumption of standardized residuals.

Table 4.15

Main Study: Residual Analysis - Rotated Component Matrix

	Component					
	1	2	3	4	5	6
SCIT-1_HHH	0.81	-0.04	0.03	-0.18	-0.18	-0.17
SCIT-12_MML	-0.67	0.13	0.23	-0.22	-0.17	-0.43
SCIT-3_HMH	0.63	-0.24	0.19	0.28	0.03	-0.17
SCIT-5_HHM	0.37	-0.14	-0.29	0.23	0.03	-0.41
SCIT-4_HMH	-0.32	-0.50	-0.57	0.04	-0.17	0.06
SCIT-7_MMM	-0.21	0.76	0.13	-0.05	-0.19	0.02
SCIT-14_LLL	-0.18	-0.17	0.31	-0.62	0.20	0.04
SCIT-2_HHH	0.13	-0.47	-0.26	0.12	-0.50	0.11
SCIT-8_MMM	-0.10	0.01	0.14	0.08	-0.11	0.84
SCIT-9_MLM	0.08	0.37	-0.32	-0.46	0.33	0.38
SCIT-11_MML	-0.08	0.66	-0.07	0.09	0.03	0.06
SCIT-6_HHM	0.00	-0.03	0.08	0.75	0.19	0.09
SCIT-10_MLM	0.00	-0.13	-0.01	0.09	0.89	-0.06
SCIT-13_LLL	-0.03	0.01	0.78	-0.01	0.00	0.17

Note. The rotation method was Varimax with Kaiser Normalization.

Scenario revisions undertaken for the main study included adding racialized representations for certain facet levels (i.e. representations that were described more often by Non-Hispanic Whites vs. other groups). The hypothesis behind this design feature was to test whether their empirical locations would be close (i.e. they represent the same level of SCI support) and whether the description for these selected facets made the scenario perceived differently by either group. One such case is scenario SCIT-8, which includes a racialized struct for the facet APE. As evidenced in Figure 4.5, SCIT-8 appears at the same difficulty level for

both racial/ethnic affiliation groups. Another example is SCIT-1, which includes a racialized struct for the facet APE. For Non-White respondents the scenario difficulty is around 1 *logit*, and at 2 *logits* for Non-Hispanic White respondents. This suggests that for this racialized struct mentioning *Black voices* and *racial injustice*, Non-Hispanic White respondents were less likely to select they were “much more proactive”.

Main Study Differential Item Functioning Analysis

To test whether the distributions of SCI are the same across groups, my second research question, I conducted a differential item functioning (DIF) test (Holland & Wainer, 1993; Millsap, 2011; Walker, 2011). In the psychological measurement realm, this technique has been used to detect measurement bias—in particular item bias in tests—in relation to student characteristics (e.g. race/ethnicity, and gender) (Millsap, 2011; Millsap & Meredith, 2007). In this dissertation, I employed DIF analysis to address my second research question about differences in the SCI-T scale as a whole, and for each scenario, across Non-Hispanic White teachers and the teachers who identified in any other racial/ethnic categories. A scenario is categorized as having DIF when teachers belonging to different racioethnic categories have the same level of SCI but differ in their probability of endorsing a given response option. For example, this analysis shows if two teachers with equal empirical levels of SCI (e.g. “moderate”) but different ethnoracial affiliation select “much more proactive” to SCIT-5 with different probabilities (e.g., Non-Hispanic White teachers select that response category less often). I checked for each scenario whether any difference in difficulty for Non-Hispanic White versus Non-White teachers could be explained by chance, using a uniform DIF analysis in WINSTEPS. Uniform DIF indicates that the DIF is in the same direction across the entire spectrum of sociocultural integration (e.g. at all levels of SCI the likelihood of selecting specific response category to a scenario is consistently higher or lower for one group than for the other) (Teresi & Fleishman, 2007). This compares to a Non-uniform DIF analysis, which would broaden the comparisons of responses of different ethnoracial groups, at all different levels of SCI.

Figure 4.8 plots the estimated difficulties for each scenario by their ethnoracial affiliation. Scenarios are shown in ascending difficulty order on the X-axis. The red line represents Non-

Hispanic White teachers, and the blue line represents Non-White teachers. Two takeaways from this figure include that the lines cross, such that the SCI-T scale is not systematically more difficult for any group; and second, there are large differences for some scenarios. Scenarios where there are no visually discernible gaps between estimated difficulties can be viewed as unlikely candidates for differential item functioning.

Figure 4.8

Main Study: Estimated Scenario Difficulties, by Ethnoracial Categories

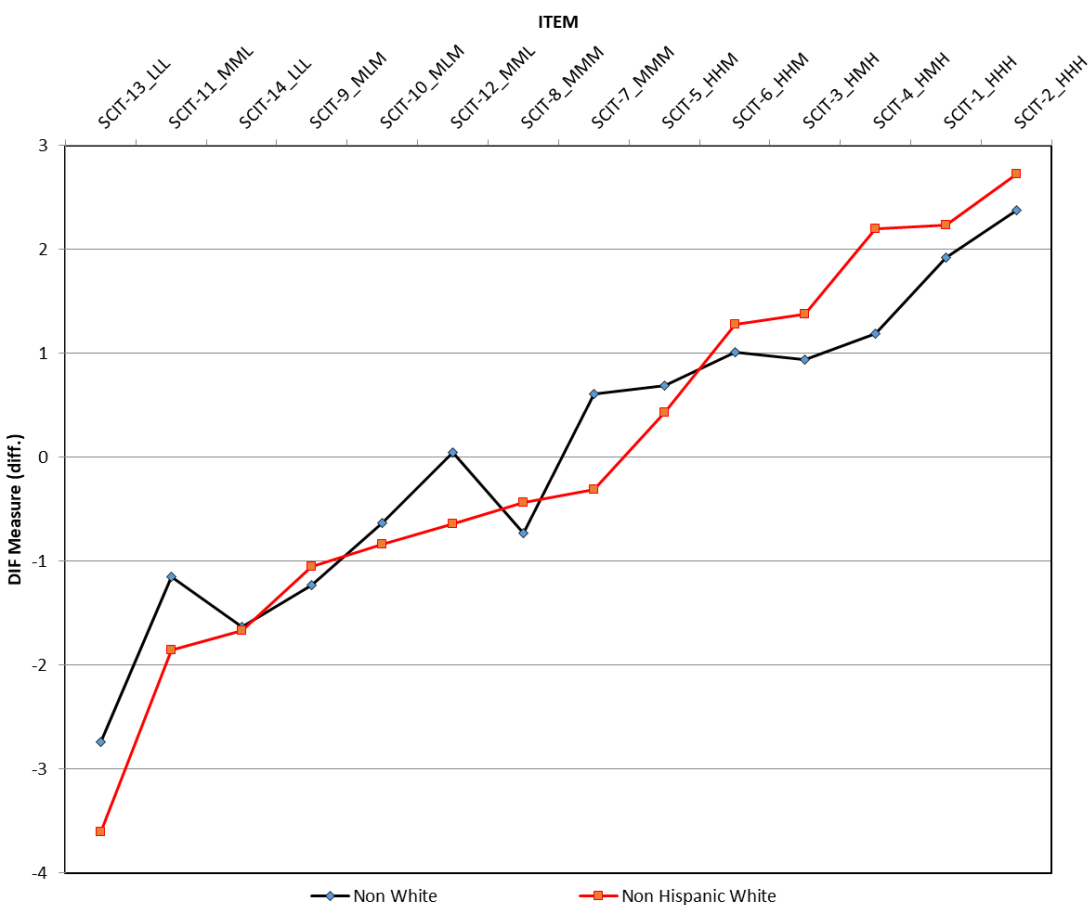


Table 4.16 provides DIF analysis results, where highlighted rows correspond to flagged scenarios. I flagged scenarios as having DIF if they obtained p-values below the 0.05 threshold for both the Rasch-Welch and Mantel tests. Using this criterion, scenario SCIT-7 was flagged for DIF. Scenarios SCIT-13, SCIT-3 and SCIT-4 were flagged for DIF with only one significance test (either the Rasch-Welch or Mantel), as seen highlighted in a different color in Table 4.16. Flagged scenarios SCIT-3 and SCIT-4 were more difficult for Non-Hispanic White teachers to say they were “much more proactive” to. The contrary is true for scenarios SCIT-7 and SCIT-13: Non-White teachers were less likely to choose being “much more proactive”. The DIF size, as indicated by the column “Size CUMLOR”, can be considered as large if compared to Zieky’s (1993, p. 342) values for the “moderate to large” category of 0.64. Table 4.17 provides the test statistics to answer whether the DIF suggested by Figure 4.8 and Table 4.16 is due to chance. Chi-square test probabilities in Table 4.17 suggest that scenarios SCIT-4 and SCIT-7 appear to have overall DIF across ethnoracial groups. SCIT-4 appears to be more difficult for Non-White teachers to select they are “much more proactive”. The situation described in this scenario does not suggest that it would be biased against any group. However, the response option wording “proactive” might have an effect similar to the pilot where teachers misunderstood the directionality. It may be that Non-White teachers were more likely to interpret “more proactive” to have a negative connotation given the situations these scenarios describe. This may be a persistent issue to be addressed in future iterations of the instrument. SCIT-7 on the other hand, has a higher difficulty estimate for Non-Hispanic White teachers. The scenario includes terms that Non-Hispanic White teachers might feel uncomfortable with if they are not at high levels of self-identification (e.g. systemic racism), which would validate the group finding it is more difficult to respond “much more proactive” than to this scenario. The discomfort or gut reaction

to the language may be consistently distracting this group from the nuances in the other facets in the scenario. Further validation could be sought using focus groups or interviews with teachers in both groups. Another option is to delete these scenarios from the scale, as there are other items that target the same facet/level combinations. This is consistent with one line of thought about the Rasch model that when items do not show desirable properties “that signifies to us not the occasion for a looser model, but the need for better items. We are looking for a core of data which can be tried as a basis for measurement because it follows our measurement specifications by conforming to our measurement model” (Wright & Masters, 1982, p. 102).

Despite DIF findings, it should be noted that Pearson correlations between scenario difficulty estimates were .96 between ethnoracial affinity groups. This suggests that the meaning of enacting sociocultural integration was invariant across Non-Hispanic White teachers and Non-White teachers. This finding provides support for the SCI model proposed in this dissertation, which places the focus on racial/ethnic and cultural identity development level instead of group affiliation. Future research to validate this model should start by providing evidence of concurrent validity against another measure of racial/ethnic or cultural identity development.

Table 4.16*Main Study: Uniform DIF Analysis, by Ethnoracial Categories*

PERSON/ CLASS	Obs- Exp Average	DIF MEASURE	DIF S.E.	PERSON/ CLASS	Obs- Exp Average	DIF MEASURE	DIF S.E.	Rasch-Welch			Mantel		Size	Name
								t	d.f.	Prob.	Chi-squ	Prob.	CUMLOR	
Non-W	0.12	1.92	0.34	W	-0.02	2.24	0.16	-0.84	26.00	0.41	2.90	0.09	-1.06	SCIT-1_HHH
Non-W	-0.33	0.61	0.35	W	0.07	-0.31	0.16	2.42	26.00	0.02	8.79	0.00	2.44	SCIT-7_MMM
Non-W	0.13	2.38	0.33	W	-0.03	2.73	0.16	-0.95	28.00	0.35	0.99	0.32	-0.64	SCIT-2_HHH
Non-W	-0.24	0.05	0.36	W	0.05	-0.64	0.16	1.74	24.00	0.09	0.76	0.38	0.49	SCIT-12_MML
Non-W	-0.10	-2.74	0.57	W	0.02	-3.61	0.29	1.37	26.00	0.18	4.68	0.03	2.25	SCIT-13_LLL
Non-W	0.16	0.94	0.35	W	-0.03	1.38	0.16	-1.13	24.00	0.27	5.24	0.02	-2.02	SCIT-3_HMH
Non-W	-0.07	-0.63	0.38	W	0.01	-0.84	0.16	0.50	23.00	0.62	0.79	0.37	0.53	SCIT-10_MLM
Non-W	0.37	1.19	0.35	W	-0.07	2.20	0.16	-2.59	24.00	0.02	3.19	0.07	-1.22	SCIT-4_HMH
Non-W	-0.09	0.69	0.36	W	0.02	0.43	0.16	0.66	23.00	0.52	0.00	0.97	-0.02	SCIT-5_HHM
Non-W	0.06	-1.23	0.40	W	-0.01	-1.05	0.16	-0.42	23.00	0.68	0.21	0.65	-0.37	SCIT-9_MLM
Non-W	-0.19	-1.15	0.39	W	0.04	-1.86	0.18	1.65	26.00	0.11	2.34	0.13	1.40	SCIT-11_MML
Non-W	-0.01	-1.63	0.41	W	0.00	-1.67	0.18	0.09	25.00	0.93	0.10	0.75	0.27	SCIT-14_LLL
Non-W	0.10	1.01	0.34	W	-0.02	1.28	0.16	-0.70	26.00	0.49	0.04	0.84	0.13	SCIT-6_HHM
Non-W	0.10	-0.73	0.37	W	-0.02	-0.43	0.16	-0.74	25.00	0.47	0.22	0.64	-0.36	SCIT-8_MMM

Table 4.17*Main Study: Uniform DIF Analysis Summary*

PERSON/ CLASS	SUMMARY CHI-SQUARED	DIF D.F.	PROB.	BETWEEN-CLASS/GROUP		ITEM	
				UNWTD MNSQ	ZSTD	Entry Number	Name
2	0.69	1	0.41	0.72	0.25	1	SCIT-1_HHH
2	5.15	1	0.02	5.86	2.17	2	SCIT-7_MMM
2	0.88	1	0.35	0.91	0.41	3	SCIT-2_HHH
2	2.78	1	0.10	3.03	1.42	4	SCIT-12_MML
2	1.77	1	0.18	1.95	1.00	5	SCIT-13_LLL
2	1.21	1	0.27	1.27	0.65	6	SCIT-3_HMH
2	0.25	1	0.62	0.26	-0.30	7	SCIT-10_MLM
2	5.81	1	0.02	6.72	2.35	8	SCIT-4_HMH
2	0.42	1	0.52	0.44	-0.04	9	SCIT-5_HHM
2	0.17	1	0.68	0.18	-0.45	10	SCIT-9_MLM
2	2.53	1	0.11	2.75	1.32	11	SCIT-11_MML
2	0.01	1	0.92	0.02	-1.12	12	SCIT-14_LLL
2	0.47	1	0.49	0.49	0.02	13	SCIT-6_HHM
2	0.53	1	0.47	0.55	0.08	14	SCIT-8_MMM

SCI-T Score Interpretation

After having determined the psychometric properties of the SCI-T, I illustrate the benefits of such a scale. In particular, I focus on the interpretability that is only possible using the RGS scale development methodology. Figure 4.9 shows the average total score on the SCI-T that aligns with SCI levels, as represented by the Y-axis (scenario/person difficulty or underlying SCI level measured in logits). The distribution of teachers is concentrated in the 36-59 total score range. There is a small group of teachers who score above 59, which represents levels of SCI higher than those captured in scenarios SCIT-2 and SCIT-1 (the highest enactments designed). There are no teachers scoring at the bottom of the distribution. While there are moderate/low and low scenarios at total scores below 36 (e.g., SCIT-11, SCIT-13), no teachers scored at this level. The nature of the construct may affect teachers' responses through social desirability bias. Future administrations should include a social desirability scale to account for this behavior.

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

Figure 4.9

Main Study: SCI-T Interpretation by Levels of Enactment

MEASURE	PERSON - MAP - ITEM
5	<more> <rare>
Average score of 67	+
	R_3q
	R_Xy
	R_2S R_31
	R_1m
4	+
Average score of 63	
	R_0x R_1d
	R_0d R_3q
	T
	R_1d R_2V R_WC
	R_10
3	+
Average score of 59	
	R_1C R_1m R_3C R_30 R_31 R_SD
	R_11 R_1n R_25 R_2C R_2f R_3H R_3M R_3s R_b9 R_tG S SCIT-2_HHH
	R_1P R_1f R_24 R_2R R_2S R_2u R_3K R_3Q R_4J
	R_12 R_1G R_1d R_1d R_1g R_1m R_1o R_24 R_2T R_2x R_33 R_3f R_3m R_Og SCIT-1_HHH
2	+
Average score of 54	R_1r R_20 R_23 R_24 R_2U R_2Y R_3I R_70 R_Tu R_us SCIT-4_HHM
	R_12 R_10 R_2D R_2t R_30 R_3d R_Xj R_Y0 R_tQ
	R_1I R_1d R_1k R_1r R_28 R_2t R_2u R_5B S
	R_1P R_1Q R_1q R_2S R_2y R_3E R_30 R_3n R_0o R_Qf R_b7 R_eb R_eg M
	R_1g R_21 R_29 R_3H R_3n R_8G R_9E R_BS R_DH SCIT-3_HMH
	R_2v R_3I R_3M R_30 R_3P R_3e R_3m SCIT-6_HHM
1	+
Average score of 48	R_1e R_1k R_1l R_1n R_1o R_2y R_3f
	R_1j R_25 R_2C R_2t R_27
	R_1J R_2T R_31 R_8e
	R_1Q R_1g R_1k R_2U S SCIT-5_HHM
	R_24 R_2f R_2w R_qR
	R_2y R_3Q
0	+M
Average score of 42	
	R_1K R_1g R_2U R_3K R_3f R_wS SCIT-7_MMM
	R_1F R_2v
	R_2z R_3o SCIT-12_MML SCIT-8_MMM
	R_28 R_An T SCIT-10_MLM
	R_2s R_3y
	R_3E R_Dl
-1	+
Average score of 36	
	SCIT-9_MLM
	SCIT-11_MML SCIT-
14_LLL	
	+
-2	+
Average score of 30	
	+
-3	+
Average score of 25	
	T SCIT-13_LLL
	+
-4	+
Average score of 21	

-5

|

 R_1m +
 <less>|<freq>

The benefits of the RGS can be further visualized in Table 4.18, which combines the score/SCI level equivalency presented in Figure 4.9, with the proposed model for teacher enactment of sociocultural integration.

Table 4.18

Main Study: Score to SCI Level Interpretation

Total Score	SCI Level	Description of Score	Selected scenarios
>59	Very High SCI	You exhibit higher levels of SCI than every scenario in the SCI-T.	
55-59	High SCI	On average, you have high levels of SID, PSE, and APE. You are “about as proactive” as SCIT-2 and SCIT-1, and “much more/more proactive” than any scenario below this level.	SCIT-2) Acceptance of self and multiculturalist view. Is not afraid to be political (e.g., speaks clearly about cultural appropriation, and ICE raids)
45-55	High/ Moderate SCI	On average, you have high levels of SID, and moderate or high levels of PSE and APE; or you have moderate levels of SID, PSE, and APE. You are “about as proactive” as SCIT-4, SCIT-3 and SCIT-6, “more proactive” than SCIT-3 and SCIT-6, or “much more/more proactive” than any scenario below this level.	SCIT-4) Acceptance of self. Engages actively in antiracist practices. SCIT-5) Ethnoracial pride. Understands systemic racism but is not comfortable exerting more effort than the administration in pursuing equity.
40-45	Moderate SCI	On average, you have moderate levels of SID, and moderate or low levels of PSE and APE. You are “about as proactive” as SCIT-7, and “much more/more proactive” than any scenario below this level, and “much less/less proactive” than any scenario above this level.	SCIT-7) Heightened awareness and group identity exploration. Is inconsistent in their use of strategies to support equitable learning.
35-40	Low/ Moderate SCI	On average, you have moderate levels of SID, and moderate or low levels of PSE and APE. You	SCIT-12) Heightened awareness of self, rejection of others or self.

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

		are "about as proactive" as SCIT-12, "much more/ proactive" than any scenario below this level, and "much less/less proactive" than any scenario above this level.	Finds being political in the classroom inappropriate.
			SCIT-9) Heightened awareness and group identity exploration. Is not comfortable leading conversations in the classroom around equity as it pertains to current events.
23-35	Low SCI	On average, you have low levels of SID, PSE, and APE. You are "about as proactive" as SCIT-13 and SCIT-14, or "much less/less proactive" than any scenario above this level.	SCIT-14) Color-blind perspective. Chooses not to be active in antiracist practices because they consider themselves unprepared/scared to say or do the wrong thing.
<23	Very Low SCI	You exhibit lower levels of SCI than every scenario in the SCI-T.	

Teachers who score in the "High" SCI category necessarily have high levels of self-identification. The model for teacher promotion of SCI assumes that the individual can continue to grow in their enactments. At this level the focus of support or development would be geared towards the breadth of enactments available at the "High" level, for example introducing the teacher to innovative ideas tried by other teachers in similar or different contexts. Encouraging creativity in enactments of SCI for all three facets may also include participation in communities of practice with this goal, as described by Scanlan and coauthors (2016). Teachers scoring in the "High/Moderate" and "Moderate/Low" categories should engage in facilitated conversations using scenarios at this level to identify the facets that require more support (e.g. is it that they do not know how else to enact agency for power equity). Support in this context means not only accompaniment in understanding, but also brainstorming about the opportunities the teacher can intentionally create to promote sociocultural integration. Finally, teachers whose scores put them in the "Low" SCI category, should engage first in professional

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

development or related activity that allows them to explore the theoretical framework of racial/ethnic and cultural identity development. At this level, the support should have a longer time frame (e.g. a school year) that allows for the reflection and conversations needed to break through the low level of SID. In general, as highlighted by Tatum (1994) and Cochran-Smith (1995) in their experiences with pre-service teachers, a one semester course exploring literature around SID is the bare minimum to meaningfully engage with the content. In addition, activities around the material and discussion require time for individuals' to digest, to put down and re-engage. The associated materials and conversations are often triggering, so the support around SID should allow for the time to reflect.

The SCI-T score interpretation table, as well as the 14 scenarios of the SCI-T, are well suited to be points of departure or alignment for existing trainings, workshops, or professional development activities with the goal of promotion sociocultural integration.

CHAPTER 5 DISCUSSION

Chapter 4 was dedicated to the analysis of data collected using two iterations of the Sociocultural Integration – Teacher scale (SCI-T). I have shown that the scale has desirable properties for the valid and reliable measurement of teachers' promotion of sociocultural integration in the K-12 context. In addition, I have shown the interpretability benefits of using the innovative Rasch/Guttman Scenario (RGS) scale development methodology (Ludlow et al., 2020) to assess non-cognitive outcomes such as SCI. This chapter summarizes and grounds the psychometric findings presented in Chapter 4, acknowledging the limitations of the study. I discuss the implications of the SCI-T as a tool for and by educators. Finally, I outline areas for future research.

Summary of Findings

I designed this dissertation to answer *can teacher promotion of sociocultural integration be measured in a psychometrically sound way, and, if so, would there be different results by teachers' racial/ethnic and cultural affiliation?* By choosing RGS methodology to design a scale to measure K-12 teachers' SCI promotion, I added two elements missing from the existing literature. First, working under the Rasch measurement framework (Rasch, 1960/80) instead of the more ubiquitous Classical Test Theory added the property of measurement invariance, which allows me to compare a teacher's SCI in the context of different administrations of the SCI-T and also to the SCI level of other teachers. This means that when the data fit the Rasch model a total score of 35 in the SCI-T has the same meaning on every occasion and for every survey taker.

The second feature of the SCI-T scale that was missing from the existing literature was the opportunity to increase survey use due to ease of interpretability.

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

Having a score interpretation table that maps total scores, scenarios, and descriptions of enactments at each SCI level is a significant innovation. This is significant not only for the community that cares about sociocultural integration but for the measurement community, as well, as it begins to embody the vision of *validity through use* – outside of cognitive outcomes. Qualitative inquiries with educators and experts, along with the pilot study, provided critical input for the SCI-T. Specifically, cognitive lab findings led me to remove names from scenarios to avoid bias, provided proof that the construct of SCI was something teachers experienced, and started conversations around what representations of SCI were authentic to teachers. More importantly, cognitive lab processes allowed me to preview how the instrument can be used in K-12 settings.

While some educators felt uncomfortable and decided to leave the cognitive lab, everyone who participated engaged in the read-aloud and shared how the scenarios resembled the experiences of them or someone they knew. This allowed peers to understand situations related to equity practice and racial/ethnic and cultural identification from the viewpoint of others and learn from their colleagues and their colleagues' experiences. It is not the job of educators of color to educate their peers about their experiences, or about their students' or communities' experiences. However, I believe it should be educational policy to provide guided opportunities for conversations around what practices for equity look like, and how such practices intersect with an individual's experiences and lenses – as framed in the model for teacher promotion of SCI.

After obtaining proof of concept, I set out to test the first version of the SCI-T. Sixty-three educators from the Two-Way Immersion Network of Catholic Schools

provided responses to the pilot study instrument. The results suggested that scenarios empirically ordered according to theory, that is, scenarios that included facets in “High” levels were more difficult for teachers to select; they were “Much more proactive” compared to scenarios in levels “Moderate” or “Low.” Questions included at the end of the survey to gather feedback, as well as interviews with educators from the sample, led to further revisions. As examples, a group of scenarios in the “Moderate” to “Low” levels had the same estimated difficulties, creating redundancy at certain SCI levels, and the highest-level scenarios had to include representations that were more difficult, as the teacher distribution had higher SCI levels than the scenario distribution. Also, at this stage, it became clear that response options had a limitation. The word “proactive” was being interpreted as a neutral adjective that, when presented in scenarios of “Low” SCI, meant being less supportive of SCI. Revisions were then submitted to experts. The main feedback revolved around the enactments of the facet of Self-Identification (SID). I made the representations more aligned with the theoretical models of racial/ethnic and cultural identity development. In addition, I included representations for all three facets that were varied in their reports of being used by different ethnoracial teacher groups. The final focus of revisions were the multiple iterations of response categories.

The COVID-19 pandemic provided a limiting backdrop for the main study. I was limited in my recruitment strategies and ability to reach K-12 teachers, and the population of interest was under tremendous stress. Adopting a snowballing strategy proved successful, allowing me to gather 163 responses for the final data collection. To limit the time spent on the survey, I did not include a social desirability scale to accompany the SCI-T. It is expected that a construct such as enactment of sociocultural integration will

induce some bias in self-reporting, which was evidenced by the distribution of teachers in both the pilot and final studies. However, a distribution of teacher scores that is higher on the SCI scale does not impede using the SCI-T as intended. The recruitment strategy also impacted how representative the sample was of the population of interest. However, it was not very different from country-level demographics by the National Center of Educational Statistics (2019). With respect to the first research question, the main study showed that the SCI-T can be a valid and reliable measure of teacher promotion of sociocultural integration. The fourteen scenarios represent a continuum of practices that support equitable opportunities to learn, anchored in teachers' identity development levels. Differential item functioning analysis provided support for the theoretical model, in that most scenarios and the overall scale behave the same across ethnoracial groups. This finding suggests that the variability is due to SCI and identity development levels – not necessarily ethnoracial affiliation, which means the answer to my second research question is no.

Study Implications

This dissertation was written in 2020, during a global racial reckoning and as COVID-19 and associated responses were disproportionately affecting people of color (CDC, 2020). The psychosocial and economic impacts of these realities on individuals and communities were only in the early stages of being quantified. A model that puts racial/ethnic and cultural identity development in the middle of teachers' promotion of sociocultural integration is thus timely – and, in my view of the literature, valid.

Development of the SCI-T contributes to the body of literature measuring non-cognitive constructs in K-12 that are related to equitable opportunities to learn.

Researchers have explored student self-esteem, self-efficacy, and identity as potential levers for cognitive outcomes. Educational frameworks to promote multicultural, culturally responsive education; social justice teaching; and sociocultural integration have been widely adopted. The responsibility to bring these frameworks to life, and live up to their potential, though, is in the hands of educators. Teachers need to be supported in their examination of the intersection of their identities and their practice. The model proposed in this dissertation for teachers' promotion of sociocultural integration acknowledges this; the model is developmental and intersectional with racial/ethnic and cultural identity. Bolstered by its innovative RGS design, which includes attractive features for score interpretability, the SCI-T provides an alternative or complement to models of teacher behavior such as Byrd's (2017) measurement of school racial climate and López's (2017b) model and instrument on teachers' critical awareness. The SCI-T also follows Chang's (2017) measurement of practice for equity in applying RGS to assess teachers' actions – rather than their knowledge or attitudes.

As author of the SCI-T, I think it is important to plainly state how I believe the instrument should be used. Tatum (1994) and Cochran-Smith's (1995) accounts of designing and leading courses for pre-service teachers about racial/ethnic and cultural identity had a considerable impact on this dissertation. I strongly support the idea that those who lead courses and conversations about equitable opportunities to learn and racial/ethnic and cultural identity are not “teaching” but facilitating. Even those with high levels of identity development acknowledge that facilitation requires being open to listening, learning, and accepting that the nature of the construct is untidy and evolutionary – not a strictly upward trajectory with a defined peak. Indeed, reactions to

the instrument and dynamics of the read-aloud exercise were primarily introspection, openness, and curiosity.

As Tatum (1994) and Cochran-Smith (1995) suggest, the perception of safety is paramount to sharing in these conversations, but the nature of the dynamic the SCI-T facilitated was as powerful as those authors report experiencing in their courses. An SCI-T-driven professional development activity is complementary to pre-service teacher courses that promote equitable opportunities to learn and/or sociocultural integration. The developmental stage of pre-service teachers who take such courses influences their experiences and practice. Recognizing this in the context of the evolutionary view of the SCI practice lens means that re-exposing oneself to racial/ethnic and cultural identity concepts and their link to a teacher's promotion of SCI is both appropriate and necessary.

The year 2020 saw an increase in interest and in suppliers of information, communities, and professional development focused on antiracism in conjunction with the Black Lives Matter movement. Trainings to sensitize educators on the topics of sociocultural integration, antiracism, and anti-bias are more widely available in numerous mediums. A tool like the SCI-T can be used and useful in the facilitation of such trainings and as a measure of impact and may thus be of interest not only to the organizers of such products, but also to education policymakers. Boosting training-experience transparency by implementing a common tool that has the desirable property of invariance could provide a platform for a nationwide conversation around SCI promotion in K-12 schooling.

Areas of Future Research

The projects on my research agenda serve to counter the perception of educational measurement as a deliberate strategy to oppress groups of students. My future research projects offer a different strategy within the quantitative paradigm: purposeful engagement in a virtuous circle of learning with practitioners, educators, and researchers involved in antiracist work. This is what educational measurement can look like when researchers of color decide what gets measured and how.

The SCI-T could be improved by changing the language of the response options. While it did not affect the main study results, in some cases scenario-level and respondent-level misfit were due to personal interpretations of the word “proactive.” Chapter 3 specifies some alternatives I considered; these could serve as a starting point for focus groups or interviews, my recommended next step to further finetune the response options.

Derived from how I define sociocultural integration and the model I proposed for teachers' promotion of SCI, SCI can be thought of as a schoolwide construct that contributes to a school's climate. A natural avenue for future research is to develop SCI scales for students and school leaders, as has happened, for example, with school climate assessment frameworks. I recommend that scales derived from the SCI-T be designed using the RGS methodological framework because of the benefits illustrated herein.

Finally, interested researchers and practitioners are encouraged to request the SCI-T enactments bank, a byproduct of the RGS scale development process. Like an item bank, the SCI enactments bank gathers representations of various actions individuals can take in support of SCI at different levels of the construct (e.g., moderate support). These

enactments are the building blocks of scenarios, the “lived experiences” that bring authenticity to this item type. I gathered enactments from interviews, focus groups, books, and existing studies, and created some with input from experts, key stakeholders, and practitioners. While numerous enactments are not included in the final scale, they hold a wealth of potential. Prospective uses for the enactment bank include activities within interventions targeting SCI, using the enactment bank as a subject for further research as individuals compare their perceived gradient of SCI with the models in a validity study, and/or turning the enactment bank into an online repository for learning/training purposes.

REFERENCES

- Acevedo-Polakovich, I. D., Cousineau, J. R., Quirk, K. M., Gerhart, J. I., Bell, K. M., & Adomako, M. S. (2014). Toward an Asset Orientation in the Study of U.S. Latina/o Youth: Biculturalism, Ethnic Identity, and Positive Youth Development Ψ . *The Counseling Psychologist, 42*(2), 201–229. <https://doi.org/10.1177/0011000013477904>
- ACLU. (2020, December 20). *What 100 Years of History Tells Us About Racism in Policing*. American Civil Liberties Union. <https://www.aclu.org/news/criminal-law-reform/what-100-years-of-history-tells-us-about-racism-in-policing/>
- Adams, J. Q., & Welsch, J. R. (1993). *Multicultural Education: Strategies for Implementation in Colleges and Universities. Volume 3*. <https://eric.ed.gov/?id=ED363211>
- Adams, V. N. (2010). *Messages in the Medium: The Relationships Among Black Media Images, Racial Identity, Body Image, and the Racial Socialization of Black Youth* [University of Pennsylvania]. <https://repository.upenn.edu/edissertations/352>
- Ahmad, F. Z., & Boser, U. (2014). America's Leaky Pipeline for Teachers of Color: Getting More Teachers of Color into the Classroom. In *Center for American Progress*. Center for American Progress. <https://eric.ed.gov/?id=ED561065>
- Aldana, A., & Byrd, C. M. (2015). School Ethnic–Racial Socialization: Learning About Race and Ethnicity Among African American Students. *The Urban Review, 47*(3), 563–576. <https://doi.org/10.1007/s11256-014-0319-0>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. American Educational Research Association.

- Anderson, G., Anderson, G. J., & Arsenault, N. (1998). *Fundamentals of educational research*. psychology press.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4), 561–573. <https://doi.org/10.1007/BF02293814>
- Andrich, D., & Marais, I. (2019). Review of Principles of Test Analysis Using Rasch Measurement Theory. In *A Course in Rasch Measurement Theory* (pp. 327–342). Springer.
- Antipkina, I., & Ludlow, L. H. (2020). Parental Involvement as a Holistic Concept Using Rasch/Guttman Scenario Scales. *Journal of Psychoeducational Assessment*, 38(7), 846–865. <https://doi.org/10.1177/0734282920903164>
- Atkinson, D. R., Morten, G., & Sue, D. W. (1979). *Counseling American minorities: A cross-cultural perspective*. Brown & Benchmark.
- Auguste, B. G., Hancock, B., & Laboissière, M. (2009). The economic cost of the US education gap. *The Mckinsey Quarterly*, June, 5.
- Awad, G. H. (2007). The Role of Racial Identity, Academic Self-Concept, and Self-Esteem in the Prediction of Academic Outcomes for African American Students. *Journal of Black Psychology*, 33(2), 188–207. <https://doi.org/10.1177/0095798407299513>
- Baez-Cruz, M., & Scanlan, M. (2016, April). Assessing School Climate In Bilingual Settings: Current Challenges and Future Directions. Paper presented at the meeting of the ,. *Knowledge to Action Achieving the Promise of Equal Educational Opportunity*. Annual Meeting, Washington, DC.
- Baglin, J. (2014). Improving your exploratory factor analysis for ordinal data: A demonstration using FACTOR. *Practical Assessment, Research, and Evaluation*, 19(1), 5.

- Baker, B. D., Farrie, D., & Sciarra, D. G. (2016). Mind the Gap: 20 Years of Progress and Retrenchment in School Funding and Achievement Gaps. *ETS Research Report Series*, 2016(1), 1–37. <https://doi.org/10.1002/ets2.12098>
- Bandura, A., & Adams, N. E. (1977). Analysis of self-efficacy theory of behavioral change. *Cognitive Therapy and Research*, 1(4), 287–310.
- Banks, J. A. (1993). The Canon Debate, Knowledge Construction, and Multicultural Education. *Educational Researcher*, 22(5), 4–14. <https://doi.org/10.3102/0013189X022005004>
- Banks, J. A. (1994). *Multicultural education: Theory and practice* (3rd ed.). Allyn and Bacon.
- Banks, J. A. (2019). *An introduction to multicultural education* (Sixth edition). Pearson.
- Bean, F. D. (2018). Growing U.S. Ethnoracial Diversity: A Positive or Negative Societal Dynamic? *The ANNALS of the American Academy of Political and Social Science*, 677(1), 229–239. <https://doi.org/10.1177/0002716218765415>
- Belle, C. (2019, January 23). *What Is Social Justice Education Anyway? (Opinion)*. Education Week. <https://www.edweek.org/teaching-learning/opinion-what-is-social-justice-education-anyway/2019/01>
- Bennett, J. (2003). *Teaching and Learning Science: A Guide to Recent Research and Its Applications*. A&C Black.
- Bennett, M. J. (1986). A developmental approach to training for intercultural sensitivity. *International Journal of Intercultural Relations*, 10(2), 179–196. [https://doi.org/10.1016/0147-1767\(86\)90005-2](https://doi.org/10.1016/0147-1767(86)90005-2)
- Bennett, M. J. (1993). Towards ethnorelativism: A developmental model of intercultural sensitivity. In R. M. Paige (Ed.), *Education for the intercultural experience* (pp. 21–71). Intercultural Press.

- Bentley, K. L., Adams, V. N., & Stevenson, H. C. (2008). Racial socialization: Roots, processes, and outcomes. In *Handbook of African American Psychology* (pp. 255–267). Thousand Oaks.
- Bertrand, M., & Mullainathan, S. (2003). *Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination* (No. w9873). National Bureau of Economic Research. <https://doi.org/10.3386/w9873>
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (Third Edition). Routledge.
- Borg, I., & Shye, S. (1995). *Facet theory*. SAGE Publications, Inc.
- Bowman, P. J., & Howard, C. (1985). Race-related Socialization, Motivation, and Academic Achievement: A Study of Black Youths in Three-Generation Families. *Journal of the American Academy of Child Psychiatry*, 24(2), 134–141. [https://doi.org/10.1016/S0002-7138\(09\)60438-6](https://doi.org/10.1016/S0002-7138(09)60438-6)
- Brisk, M. E. (2006). *Bilingual Education: From Compensatory To Quality Schooling*. Routledge.
- Bristol, T. J., & Martin-Fernandez, J. (2019). The Added Value of Latinx and Black Teachers for Latinx and Black Students: Implications for Policy. *Policy Insights from the Behavioral and Brain Sciences*, 6(2), 147–153. <https://doi.org/10.1177/2372732219862573>
- Broman, C. L., Neighbors, H. W., & Jackson, J. S. (1988). Racial Group Identification Among Black Adults*. *Social Forces*, 67(1), 146–158. <https://doi.org/10.1093/sf/67.1.146>
- Bronfenbrenner, U. (1979). *The Ecology of Human Development: Experiments by Nature and Design*. Harvard University Press.

- Brown, T. L., & Krishnakumar, A. (2007). Development and Validation of the Adolescent Racial and Ethnic Socialization Scale (ARESS) in African American Families. *Journal of Youth and Adolescence*, 36(8), 1072–1085. <https://doi.org/10.1007/s10964-007-9197-z>
- Buchanan, T. (2000). Potential of the Internet for personality research. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 121–140). Academic Press.
- Byrd, C. M. (2017). The complexity of school racial climate: Reliability and validity of a new measure for secondary students. *British Journal of Educational Psychology*, 87(4), 700–721. <https://doi.org/10.1111/bjep.12179>
- Canter, D. V. (1996). *The potential of facet theory for applied social psychology*. Dartmouth Publishing Company.
- Carter, R. T. (1991). Racial Identity Attitudes and Psychological Functioning. *Journal of Multicultural Counseling and Development*, 19(3), 105–114. <https://doi.org/10.1002/j.2161-1912.1991.tb00547.x>
- Carter, R. T., & Goodwin, A. L. (1994). Chapter 7: Racial Identity and Education. *Review of Research in Education*, 20(1), 291–336. <https://doi.org/10.3102/0091732X020001291>
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2), 245–276.
- CDC. (2020, February 11). *Community, Work, and School*. Centers for Disease Control and Prevention. <https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/race-ethnicity.html>
- Chang (2017).pdf. (n.d.).
- Chang, A., Martin, R., & Marrapodi, E. (2020, August 16). *Summer Of Racial Reckoning*. NPR. <https://www.npr.org/2020/08/16/902179773/summer-of-racial-reckoning-the-match-lit>

- Chang, W.-C. C. (2017). Measuring the complexity of teachers' enactment of practice for equity: A Rasch model and facet theory-based approach. *ProQuest LLC*.
- Cochran-Smith, M. (1995). Uncertain Allies: Understanding the Boundaries of Race and Teaching. *Harvard Educational Review*, 65(4), 541–571.
<https://doi.org/10.17763/haer.65.4.m77263886583326v>
- Cochran-Smith, M. (1999). Learning to teach for social justice. In G. A. Griffin & M. Early (Eds.), *The education of teachers: Ninety-eighth yearbook of the national society for the study of education* (pp. 114–144). University of Chicago Press.
- Cochran-Smith, M. (2010). Toward a Theory of Teacher Education for Social Justice. In A. Hargreaves, A. Lieberman, M. Fullan, & D. Hopkins (Eds.), *Second International Handbook of Educational Change* (pp. 445–467). Springer Netherlands.
https://doi.org/10.1007/978-90-481-2660-6_27
- Cochran-Smith, M., Ell, F., Grudnoff, L., Haigh, M., Hill, M., & Ludlow, L. (2016). Initial teacher education: What does it take to put equity at the center? *Teaching and Teacher Education*, 57, 67–78. <https://doi.org/10.1016/j.tate.2016.03.006>
- Colby, S. L., & Ortman, J. M. (2015). *Projections of the Size and Composition of the U.S. Population: 2014 to 2060* (No. P25-1143; Current Population Reports, p. 13). U.S. Census Bureau.
- Conchas, G. (2001). Structuring Failure and Success: Understanding the Variability in Latino School Engagement. *Harvard Educational Review*, 71(3), 475–505.
<https://doi.org/10.17763/haer.71.3.280w814v1603473k>

- Constantine, M. G., & Blackmon, S. M. (2002). Black Adolescents' Racial Socialization Experiences: Their Relations to Home, School, and Peer Self-Esteem. *Journal of Black Studies*, 32(3), 322–335. <https://doi.org/10.1177/002193470203200303>
- Cox, D. R., & Snell, E. J. (1989). *Analysis of Binary Data, Second Edition*. CRC Press.
- Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Holt, Rinehart and Winston, 6277 Sea Harbor Drive, Orlando, FL 32887 (\$44).
- Cross, W.E., Jr. (1991). *Shades of Black: Diversity in African-American identity*. Temple University Press.
- Cross, William E. (1971). The Negro-to-Black Conversion Experience. *Black World*.
- Cross, William E., & Cross, T. B. (2008). Theory, Research, and Models. In S. M. Quintana & C. McKown (Eds.), *Handbook of Race, Racism, and the Developing Child* (pp. 154–181). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118269930.ch8>
- Cross, William E., & Vandiver, B. J. (2001). Nigrescence theory and measurement. In C. M. Alexander, J. M. Casas, J. G. Ponterotto, & L. A. Suzuki (Eds.), *Handbook of multicultural counseling* (Vol. 2, pp. 371–393). Sage Thousand Oaks, CA.
- D'Andrea, M., Daniels, J., & Heck, R. (1991). Evaluating the impact of multicultural counseling training. *Journal of Counseling & Development*, 70(1), 143–150.
- Darling-Hammond, L. (2007). Race, inequality and educational accountability: The irony of 'No Child Left Behind.' *Race Ethnicity and Education*, 10(3), 245–260.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. The Guilford Press.
- de Brey, C., Musu, L., McFarland, J., Wilkinson-Flicker, S., Diliberti, M., Zhang, A., Branstetter, C., & Wang, X. (2019). *Status and Trends in the Education of Racial and*

- Ethnic Groups 2018* (NCES 2019-038). National Center for Education Statistics.
<https://nces.ed.gov/pubs2019/2019038.pdf>
- De Jong, E. J. (2011). *Foundations for multilingualism in education: From principles to practice*. Caslon Pub.
- Der-Karabetian, A. (1980). Relation of two cultural identities of Armenian-Americans. *Psychological Reports*, 47(1), 123–128.
- DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Sage Publications, Inc.
- Dooly, M. (2005). How Aware are They? Research into Teachers' Attitudes about Linguistic Diversity. *Language Awareness*, 14(2–3), 97–111.
<https://doi.org/10.1080/09658410508668827>
- Enterline, S., Cochran-Smith, M., Ludlow, L. H., & Mitescu, E. (2008). Learning to Teach for Social Justice: Measuring Change in the Beliefs of Teacher Candidates. *The New Educator*, 4(4), 267–290. <https://doi.org/10.1080/15476880802430361>
- Erikson, E. H. (1963). *Childhood and society* (2d ed., rev.enl.). Norton.
- Erikson, E. H. (1997). *The life cycle completed: A review* (Extended version / with new chapters by Joan M. Erikson.). WWNorton.
- Feinauer, E., & Howard, E. R. (2014). Attending to the third goal: Cross-cultural competence and identity development in two-way immersion programs. *Journal of Immersion and Content-Based Language Education*, 2(2), 257–272.
- Fisher, A. E., Fisher, S., Arsenault, C., Jacob, R., & Barnes-Najor, J. (2020). The Moderating Role of Ethnic Identity on the Relationship Between School Climate and Self-Esteem for

- African American Adolescents. *School Psychology Review*, 49(3), 291–305.
<https://doi.org/10.1080/2372966X.2020.1760690>
- Furlong, M. J., O'brennan, L. M., & You, S. (2011). Psychometric properties of the Add Health School Connectedness Scale for 18 sociocultural groups. *Psychology in the Schools*, 48(10), 986–997. <https://doi.org/10.1002/pits.20609>
- Gallegos, P. V., & Ferdman, B. M. (2007). Identity Orientations of Latinos in the United States: *The Business Journal of Hispanic Research*, 1(1), 26–41.
- García, S. B., & Ortiz, A. A. (2006). Preventing Disproportionate Representation: Culturally and Linguistically Responsive Prereferral Interventions. *Teaching Exceptional Children*, 38(4), 64–68. <https://doi.org/10.1177/004005990603800410>
- García-Nevarez, A. G., Stafford, M. E., & Arias, B. (2005). Arizona Elementary Teachers' Attitudes Toward English Language Learners and the Use of Spanish in Classroom Instruction. *Bilingual Research Journal*, 29(2), 295–317.
<https://doi.org/10.1080/15235882.2005.10162837>
- Gay, G. (2000). *Culturally responsive teaching theory, research, and practice*. Teachers College Press.
- Gay, G. (2002). Preparing for culturally responsive teaching. *Journal of Teacher Education*, 53(2), 106–116.
- Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student–teacher demographic match on teacher expectations. *Economics of Education Review*, 52, 209–224. <https://doi.org/10.1016/j.econedurev.2016.03.002>

- Gillette, M., & Boyle-Baise, M. (1996). Multicultural Education at the Graduate Level: Assisting Teachers in Gaining Multicultural Understandings. *Theory and Research in Social Education, 24*(3), 273–293.
- Gomez, A., Grantham-Philips, W., Hughes, T., Jervis, R., Plevin, R., Stanley, K., Wagner, D., della Cava, M., Barfield Berry, D., & Nichols, M. (2020, October 21). “An unbelievable chain of oppression”: America’s history of racism was a preexisting condition for COVID-19. USA Today. <https://www.usatoday.com/in-depth/news/nation/2020/10/12/coronavirus-deaths-reveal-systemic-racism-united-states/5770952002/>
- Goode, T. D. (2010). *A Guide for Using the Cultural and Linguistic Competence Family Organizational Assessment Instrument*. (p. 28). National Center for Cultural Competence, Georgetown University Center for Child and Human Development.
- Grant, C. A., & Sleeter, C. E. (2006). Turning on Learning: Five Approaches for Multicultural Teaching Plans for Race, Class, Gender and Disability, 4th Edition. In *Jossey-Bass, An Imprint of Wiley*. Jossey-Bass, An Imprint of Wiley.
- Green, C., Kalvaitis, D., & Worster, A. (2016). Recontextualizing psychosocial development in young children: A model of environmental identity development. *Environmental Education Research, 22*(7), 1025–1048. <https://doi.org/10.1080/13504622.2015.1072136>
- Guardiola Castillo, I. V. (2014). *Teachers’ and school counselors’ perceptions of their cultural competence in working with newly arrived Latino immigrant students: A mixed methods research study* [Ph.D., Sam Houston State University]. <http://search.proquest.com/docview/1544144528/abstract/2FB975DEF2A34A35PQ/1>

- Guttman, L. (1954). An outline of some new methodology for social research. *Public Opinion Quarterly*, 18(4), 395–404.
- Guttman, L. (1959). A Structural Theory For Intergroup Beliefs and Action. *American Sociological Review*, 24(3), 318–328.
- Hackett, P. M. W. (2014). A Facet Theory Model for Integrating Contextual and Personal Experiences of International Students. *Journal of International Students*, 4(2), 163–176.
<https://doi.org/10.32674/jis.v4i2.476>
- Hackett, P. M. W. (2016). Facet Theory and the Mapping Sentence As Hermeneutically Consistent Structured Meta-Ontology and Structured Meta-Mereology. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.00471>
- Helms, J. E. (1984). Toward a Theoretical Explanation of the Effects of Race on Counseling A Black and White Model. *The Counseling Psychologist*, 12(4), 153–165.
<https://doi.org/10.1177/0011000084124013>
- Helms, J. E. (1990). *Black and White racial identity: Theory, research, and practice* (pp. xviii, 262). Greenwood Press.
- Helms, J. E., & Carter, R. T. (1991). Relationships of White and Black racial identity attitudes and demographic similarity to counselor preferences. *Journal of Counseling Psychology*, 38(4), 446–457. <https://doi.org/10.1037/0022-0167.38.4.446>
- Hill, S. A. (2001). Class, race, and gender: Dimensions of child rearing in African American families. *Journal of Black Studies*, 31(4), 494–508.
- Ho, M.-J., & Lee, K.-L. (2007). Reliability and validity of three cultural competency measures. *Medical Education*, 41(5), 519–519.

- Hochschild, J. L., & Weaver, V. (2007). The Skin Color Paradox and the American Racial Order. *Social Forces*, *86*(2), 643–670. <https://doi.org/10.1093/sf/86.2.643>
- Holland, P. W., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Lawrence Erlbaum Associates.
- Horse, P. G. (2005). Native American identity. *New Directions for Student Services*, *109*, 61–68.
- Howard, E. R., Sugarman, J., Christian, D., Lindholm-Leary, K. J., & Rogers, D. (2007). *Guiding principles for dual language education*. Center for Applied Linguistics Washington, DC.
- Hsieh, T., Liu, Y., & Simpkins, S. D. (2019). Changes in United States Latino/a High School Students' Science Motivational Beliefs: Within Group Differences Across Science Subjects, Gender, Immigrant Status, and Perceived Support. *Frontiers in Psychology*, *10*. <https://doi.org/10.3389/fpsyg.2019.00380>
- Hughes, D. (2003). Correlates of African American and Latino Parents' Messages to Children About Ethnicity and Race: A Comparative Study of Racial Socialization. *American Journal of Community Psychology*, *31*(1–2), 15–33. <https://doi.org/10.1023/A:1023066418688>
- Iramaneerat, C., Smith Jr., E. V., & Smith, R. M. (2008). An Introduction to Rasch Measurement. In J. Osborne (Ed.), *Best Practices in Quantitative Methods* (pp. 50–70). SAGE Publications, Inc.
- Jacob, B., & Ludwig, J. (2008). *Improving educational outcomes for poor children*. National Bureau of Economic Research.
- Kaiser, H. F. (1960). Directional statistical decisions. *Psychological Review*, *67*(3), 160.

Kardong-Edgren, S., & Campinha-Bacote, J. (2008). Cultural competency of graduating US Bachelor of Science nursing students. *Contemporary Nurse*, 28(1–2), 37–44.

Kearney, M. S. (2019, June 7). To thrive, American children need a stronger safety net. *Brookings*. <https://www.brookings.edu/opinions/to-thrive-american-children-need-a-stronger-safety-net/>

Kendi, I. X. (2019). *How to Be an Antiracist*. Random House Publishing Group.
<https://books.google.com/books?id=6pNbDwAAQBAJ>

Kim, J. (1981). *Processes of Asian American identity development: A study of Japanese American women's perceptions of their struggle to achieve positive identities as Americans of Asian ancestry*. University of Massachusetts Amherst.

Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2020). Projecting the potential impacts of COVID-19 school closures on academic achievement. In *EdWorkingPapers.com*. Annenberg Institute at Brown University.
<https://www.edworkingpapers.com/ai20-226>

Ladson-Billings, G. (1995). Toward a Theory of Culturally Relevant Pedagogy. *American Educational Research Journal*, 32(3), 465–491.
<https://doi.org/10.3102/00028312032003465>

Ladson-Billings, G. (2006). From the Achievement Gap to the Education Debt: Understanding Achievement in U.S. Schools. *Educational Researcher*, 35(7), 3–12.
<https://doi.org/10.3102/0013189X035007003>

Ladson-Billings, G. (2009). *The dreamkeepers: Successful teachers of African American children*. John Wiley & Sons.

- Lee, E. (1997). Anti-racist education: Pulling Together to Close the Gaps. In E. Lee, D. Menkart, & M. Okazawa-Rey (Eds.), *Beyond Heroes and Holidays: A Practical Guide to K-12 Anti-Racist, Multicultural Education and Staff Development*. Network of Educators on the Americas.
- Lesane-Brown, C. L., Brown, T. N., Caldwell, C. H., & Sellers, R. M. (2005). The Comprehensive Race Socialization Inventory. *Journal of Black Studies*, 36(2), 163–190.
- Levine, R., Huberman, M., & Buckner, K. (2002). *The Measurement of Instructional Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Grade Students and Teachers to Questionnaire Items*. Working Paper Series (p. 296). National Center for Education Statistics. <https://eric.ed.gov/?id=ED469768>
- Linacre, J. M. (2016). *Winsteps® Rasch measurement computer program User's Guide*. Winsteps.
- Linacre, J. M. (2020). *WINSTEPS* (4.8.0) [Computer software]. Winsteps.com. Available from: <https://www.winsteps.com/winpass.htm>
- Lipsky, M. (1980). *Street-Level Bureaucracy, 30th Anniversary Edition: Dilemmas of the Individual in Public Service*. Russell Sage Foundation.
- Lissitz, R. W. (Ed.). (2009). *The concept of validity: Revisions, new directions, and applications*. Information Age Publishing, INC.
- Loftin, C., Hartin, V., Branson, M., & Reyes, H. (2013). Measures of Cultural Competence in Nurses: An Integrative Review. *The Scientific World Journal*, 2013, e289101. <https://doi.org/10.1155/2013/289101>

- López, F. A. (2017a). Altering the Trajectory of the Self-Fulfilling Prophecy: Asset-Based Pedagogy and Classroom Dynamics. *Journal of Teacher Education*, 68(2), 193–212.
<https://doi.org/10.1177/0022487116685751>
- López, F. A. (2017b). Altering the Trajectory of the Self-Fulfilling Prophecy: Asset-Based Pedagogy and Classroom Dynamics. *Journal of Teacher Education*, 68(2), 193–212.
<https://doi.org/10.1177/0022487116685751>
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.
- Lucas, T., & Villegas, A. M. (2013). Preparing Linguistically Responsive Teachers: Laying the Foundation in Preservice Teacher Education. *Theory Into Practice*, 52(2), 98–109.
<https://doi.org/10.1080/00405841.2013.770327>
- Ludlow, L. H. (1983). *The analysis of Rasch model residuals* [PhD Thesis]. University of Chicago, Department of Education.
- Ludlow, L. H., Baez-Cruz, M., Chang, W.-C. C., & Reynolds, K. (2020). Rasch/Guttman Scenario (RGS) Scales: A Methodological Framework. *Journal of Applied Measurement*, 21(4), 361–378.
- Ludlow, L. H., & Haley, S. M. (1995). Rasch model logits: Interpretation, use, and transformation. *Educational and Psychological Measurement*, 55(6), 967–975.
- Ludlow, L. H., Matz-Costa, C., Johnson, C., Brown, M., Besen, E., & James, J. B. (2014). Measuring Engagement in Later Life Activities: Rasch-Based Scenario Scales for Work, Caregiving, Informal Helping, and Volunteering. *Measurement and Evaluation in Counseling and Development*, 47(2), 127–149.

- Ludlow, L. H., Matz-Costa, C., & Klein, K. (2019). Enhancement and Validation of the Productive Engagement Portfolio–Scenario (PEP–S8) Scales. *Measurement and Evaluation in Counseling and Development*, 52(1), 15–37.
- Ludlow, L. H., Reynolds, K. R., Baez-Cruz, M., & Chang, W.-C. C. (In Press). Enhancing the interpretation of scores through Rasch based scenario-style items. In G. Harbaugh & U. Luhanga (Eds.), *Basic Elements of Survey Research in Education: Addressing the Problems Your Advisor Never Told You About*.
- Marais, I., & Andrich, D. (2005). *Studies on the effect of violations of local independence on scale in Rasch models*. (No. 1; ARC Linkage Research Report). Murdoch University.
- Marshall, S. (1995). Ethnic socialization of African American children: Implications for parenting, identity development, and academic achievement. *Journal of Youth and Adolescence*, 24(4), 377–396. <https://doi.org/10.1007/BF01537187>
- Martin, A. D., & Strom, K. J. (2016). Toward a Linguistically Responsive Teacher Identity: An Empirical Review of the Literature. *International Multilingual Research Journal*, 10(4), 239–253. <https://doi.org/10.1080/19313152.2016.1189799>
- McAdoo, H. P. (1985). Racial attitudes and self-concept of young Black children over time. In *Black children: Social, educational, and parental environments* (pp. 213–242). Sage Publications, Inc.
- McAllister, G., & Irvine, J. J. (2000). Cross Cultural Competency and Multicultural Teacher Education. *Review of Educational Research*, 70(1), 3–24. <https://doi.org/10.3102/00346543070001003>
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437.

Mezirow, J. (1978). Perspective Transformation. *Adult Education*, 28(2), 100–110.

<https://doi.org/10.1177/074171367802800202>

Miller, L. (2015, May 19). Can Racism Be Stopped in the Third Grade? *The Cut*.

<https://www.thecut.com/2015/05/can-fieldston-un-teach-racism.html>

Millsap, R. E. (2011). *Statistical Approaches to Measurement Invariance*. Routledge.

Millsap, R. E., & Meredith, W. (2007). Factor analysis models as approximation. In R. Cudeck & R. C. MacCallum, *Factor analysis at 100: Historical developments and future directions*. Lawrence Erlbaum Associates.

Miramontes, O. B., Nadeau, A., & Commins, N. L. (2011). Restructuring Schools for Linguistic Diversity: Linking Decision Making to Effective Programs. Second Edition. Language & Literacy Series. In *Teachers College Press*. Teachers College Press.

Mitchum, N. T. (1989). Increasing self-esteem in native-american children. *Elementary School Guidance & Counseling*, 23(4), 266–271.

Mokkan, R. J., & Lewis, C. (1982). A Nonparametric Approach to the Analysis of Dichotomous Item Responses. *Applied Psychological Measurement*, 6(4), 417–430.

<https://doi.org/10.1177/014662168200600404>

National Education Association. (2015). Understanding the gaps: Who are we leaving behind and how far. *NEA Education Policy and Practice & Priority Schools Departments and Center for Great Public Schools*.

NCES. (2019). *The Condition of Education* (Characteristics of Public School Teachers, p. 7).

https://nces.ed.gov/programs/coe/indicator_clr.asp

Nieto, S., & Rolón, C. (1995, November). *Preparation and professional development of teachers: A perspective from two Latinas*. Defining the Knowledge Base for Urban

Teacher Education, Atlanta GA.

https://scholar.google.com/scholar_lookup?hl=en&publication_year=1995&conference=Paper+presented+at+the+invitational+conference+on+Defining+the+Knowledge+Base+for+Urban+Teacher+Education&author=S+Nieto&author=C+Rolon&title=The+preparation+and+professional+development+of+teachers%3A+A+perspective+from+two+Latinas

Nieto, Sonia. (1992). *Affirming diversity: The sociopolitical context of multicultural education*.

ERIC.

Nieto, Sonia. (1994). Moving Beyond Tolerance in Multicultural Education. *Multicultural Education*, 8.

Nieto, Sonia. (1999). What Does It Mean To Affirm Diversity?. *School Administrator*, 56(6), 32–35.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, 32(3), 396–402.

Parham, T. A., & Helms, J. E. (1985a). Attitudes of racial identity and self-esteem of Black students: An exploratory investigation. *Journal of College Student Personnel*.

Parham, T. A., & Helms, J. E. (1985b). Relation of racial identity attitudes to self-actualization and affective states of Black students. *Journal of Counseling Psychology*, 32(3), 431–440. <https://doi.org/10.1037/0022-0167.32.3.431>

Pedersen, D. M. (1999). Validating a centrality model of self-identity. *Social Behavior and Personality*, 27(1), 73. <https://doi.org/10.2224/sbp.1999.27.1.73>

- Peguero, A. A., & Bondy, J. M. (2011). Immigration and students' relationship with teachers. *Education and Urban Society, 43*(2), 165–183.
<https://doi.org/10.1177/0013124510380233>
- Peña, E., & Bedore, L. (2010). Bilingualism. In *Encyclopedia of Cross-Cultural School Psychology* (Vol. 2, pp. 153–164). Springer.
- Perng, S.-J., & Watson, R. (2012). Construct validation of the nurse cultural competence scale: A hierarchy of abilities. *Journal of Clinical Nursing, 21*(11–12), 1678–1684.
- Phinney, J. S. (2000). Ethnic and racial identity: Ethnic identity. In *Encyclopedia of Psychology, Vol. 3*. (pp. 254–259). Oxford University Press. <https://doi.org/10.1037/10518-122>
- Phinney, J. S., & Rotheram, M. J. (1986). *Children's Ethnic Socialization: Pluralism and Development*. SAGE Publications, Inc.
- Pohan, C. A., & Aguilar, T. E. (2001). Measuring educators' beliefs about diversity in personal and professional contexts. *American Educational Research Journal, 38*(1), 159–182.
- Potts, R. G. (2003). Emancipatory Education Versus School-Based Prevention in African American Communities. *American Journal of Community Psychology, 31*(1), 173–183.
<https://doi.org/10.1023/A:1023039007302>
- QSR International. (1999). *NVivo Qualitative Data Analysis Software*. Available from <https://qsrinternational.com/nvivo/nvivo-products/>
- Quintana, S. M., & McKown, C. (2008). *Handbook of Race, Racism, and the Developing Child*. John Wiley & Sons.
- Quiocho, A., & Rios, F. (2000). The Power of Their Presence: Minority Group Teachers and Schooling. *Review of Educational Research, 70*(4), 485–528.
<https://doi.org/10.3102/00346543070004485>

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests* ((Expanded edition, 1980. Chicago, IL: University of Chicago Press.)). Danish Institute for Educational Research.
- Rew, L., Becker, H., Cookston, J., Khosropour, S., & Martinez, S. (2003). Measuring cultural awareness in nursing students. *Journal of Nursing Education, 42*(6), 249–257.
- Reynolds, K. A. (2020). *Measuring Students' Perceptions of Faculty Availability Outside of Class Using Rasch/Guttman Scenario Scales* [PhD Thesis]. Boston College. Lynch School of Education.
- Rossi, P. H., & Anderson, A. B. (1982). The factorial survey approach: An introduction. In P. Rossi & S. L. Nock (Eds.), *Measuring Social Judgments* (pp. 1–25). SAGE Publications, Inc.
- Sadler, G. R., Lee, H.-C., Seung-Hwan Lim, R., & Fullerton, J. (2010). Recruiting hard-to-reach United States population sub-groups via adaptations of snowball sampling strategy. *Nursing & Health Sciences, 12*(3), 369–374. <https://doi.org/10.1111/j.1442-2018.2010.00541.x>
- Scanlan, M., Kim, M., Burns, M. B., & Vuilleumier, C. (2016). Poco a Poco: Leadership Practices Supporting Productive Communities of Practice in Schools Serving the New Mainstream. *Educational Administration Quarterly, 52*(1), 3–44. <https://doi.org/10.1177/0013161X15615390>
- Scanlan, M., & López, F. A. (2014). *Leadership for culturally and linguistically responsive schools*. Routledge.
- Scarr, S., & Salapatek, P. (1973). *Socialization*. Columbus, Ohio, C.E. Merrill Pub. Co. <http://archive.org/details/socialization0000scar>

- Scott Jr, L. D. (2003). The relation of racial identity and racial socialization to coping with discrimination among African American adolescents. *Journal of Black Studies*, 33(4), 520–538.
- Shye, S. (Ed.). (1978). *Theory construction and data analysis in the behavioral sciences*. Jossey-Bass.
- Shye, S. (1981). An integrated method for scaling subjects and structuring their multivariate attitudes: Description and illustration of partial order scalogram and lattice analysis. *Jerusalem: The Israeli Institute of Applied Social Research*.
- Shye, Samuel. (1988). Inductive and deductive reasoning: A structural reanalysis of ability tests. *Journal of Applied Psychology*, 73(2), 308.
- Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge University Press.
- Singleton, C. S. (2009). *An examination of the construct validity of the White Racial Identity Attitudes Scale (WRIAS) and the Oklahoma Racial Attitudes Scale (ORAS)* [ProQuest Dissertations Publishing]. <https://search.proquest.com/docview/305072983?pq-origsite=summon>
- Siwatu, K. O. (2005). *Exploring the factors that influence preservice teachers' culturally responsive teaching self-efficacy and outcome expectancy beliefs* [Ph.D., The University of Nebraska - Lincoln]. <https://search.proquest.com/docview/305426343/abstract/21403B8D94224A7APQ/1>
- Siwatu, K. O. (2007). Preservice teachers' culturally responsive teaching self-efficacy and outcome expectancy beliefs. *Teaching and Teacher Education*, 23(7), 1086–1101.

- Sleeter, C. (2013). Teaching for Social Justice in Multicultural Classrooms. *Multicultural Education Review*, 5(2), 1–19. <https://doi.org/10.1080/2005615X.2013.11102900>
- Sleeter, C. E. (1992). Resisting Racial Awareness: How Teachers Understand the Social Order from Their Racial, Gender, and Social Class Locations. *The Journal of Educational Foundations*, 6(2), 7–32.
- Sleeter, C., & Grant, C. (1987). An Analysis of Multicultural Education in the United States. *Harvard Educational Review*, 57(4), 421–445.
<https://doi.org/10.17763/haer.57.4.v810xr0v3224x316>
- Smith, E. P., Atkins, J., & Connell, C. M. (2003). Family, School, and Community Factors and Relationships to Racial–Ethnic Attitudes and Academic Achievement. *American Journal of Community Psychology*, 32(1–2), 159–173. <https://doi.org/10.1023/A:1025663311100>
- Smith, E. V., Conrad, K. M., Chang, K., & Piazza, J. (2002). An introduction to Rasch measurement for scale development and person assessment. *Journal of Nursing Measurement*, 10(3), 189–206.
- Smith Jr., E. V. (2004b). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. In E. V. Smith Jr. & R. M. Smith (Eds.), *Introduction to Rasch measurement* (pp. 93–122). JAM Press.
- Smith, R. M., Linacre, J. M., & Smith, E. V. (2003). Guidelines for manuscripts. *Journal of Applied Measurement*, 4(2), 198–204.
- Southern Poverty Law Center. (2021). *Learning for Justice*. Learning for Justice Education Resources. <https://www.learningforjustice.org/home>
- Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science, New Series*, 103(2684), 677–680.

- Stevenson, H. C. (1994). Validation of the Scale of Racial Socialization for African American Adolescents: Steps toward Multidimensionality. *Journal of Black Psychology, 20*(4), 445–468. <https://doi.org/10.1177/00957984940204005>
- Suárez-Orozco, C., & Suárez-Orozco, M. M. (2001). *Children of immigration* (pp. x, 206). Harvard University Press.
- Taggart, A. (2018). Latina/o Students in K-12 Schools: A Synthesis of Empirical Research on Factors Influencing Academic Achievement. *Hispanic Journal of Behavioral Sciences, 40*(4), 448–471. <https://doi.org/10.1177/0739986318793810>
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The Social Psychology of Intergroup Relations* (pp. 33–47). Brooks/Cole.
- Tatum, B. D. (1994). Teaching white students about racism: The search for white allies and the restoration of hope. *Teachers College Record (1970), 95*(4), 462-.
- Taylor, I. (2007). Discretion and Control in Education: The Teacher as Street-level Bureaucrat. *Educational Management Administration & Leadership, 35*(4), 555–572. <https://doi.org/10.1177/1741143207081063>
- Taylor, R. J. (1986). Receipt of support from family among Black Americans: Demographic and familial differences. *Journal of Marriage and the Family, 67*–77.
- Teresi, J. A., & Fleishman, J. A. (2007). Differential item functioning and health assessment. *Quality of Life Research, 16*(S1), 33–42. <https://doi.org/10.1007/s11136-007-9184-6>
- Tesser, A. (1988). Toward a Self-Evaluation Maintenance Model of Social Behavior. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 21, pp. 181–227). Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)60227-0](https://doi.org/10.1016/S0065-2601(08)60227-0)

- Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A Review of School Climate Research. *Review of Educational Research, 83*(3), 357–385.
<https://doi.org/10.3102/0034654313483907>
- Tomlinson, L. M. (1996). *Applying Banks' Typology of Ethnic Identity Development and Curriculum Goals to Story Content, Classroom Discussion, and the Ecology of Classroom and Community: Phase One. Instructional Resource No. 24.*
<https://eric.ed.gov/?id=ED395308>
- Turner, J. C., & Onorato, R. S. (2014). Social Identity, Personality, and the Self-Concept: A Self-Categorization Perspective. In T. R. Tyler, R. M. Kramer, & O. P. John (Eds.), *The Psychology of the Social Self* (p. 9781317778288). Psychology Press.
- Umaña-Taylor, A. J., Yazedjian, A., & Bámaca-Gómez, M. (2004). Developing the Ethnic Identity Scale Using Eriksonian and Social Identity Perspectives. *Identity, 4*(1), 9–38.
https://doi.org/10.1207/S1532706XID0401_2
- U.S. Department of Education. (2016, March 16). *Department of Education Releases Resources on Improving School Climate.* Press Release. <https://www.ed.gov/news/press-releases/department-education-releases-resources-improving-school-climate>
- Vandiver, B. J., Cross, W. E., Worrell, F. C., & Fhagen-Smith, P. E. (2002). Validating the Cross Racial Identity Scale. *Journal of Counseling Psychology, 49*(1), 71–85.
<https://doi.org/10.1037/0022-0167.49.1.71>
- Vandiver, B. J., Fhagen-Smith, P. E., Cokley, K. O., Cross, W. E., & Worrell, F. C. (2001). Cross's Nigrescence Model: From Theory to Scale to Theory. *Journal of Multicultural Counseling and Development, 29*(3), 174–200. <https://doi.org/10.1002/j.2161-1912.2001.tb00516.x>

- Villegas, A. M. (1991). *Culturally Responsive Pedagogy for the 1990s and Beyond. Trends and Issues Paper No. 6*. ERIC Clearinghouse on Teacher Education, One Dupont Circle, NW, Suite 610, Washington, DC 20036-1186 (\$12. <https://eric.ed.gov/?id=ED339698>)
- Villegas, A. M., & Lucas, T. (2002). *Educating culturally responsive teachers: A coherent approach*. Suny Press.
- Voight, A., Austin, G., & Hanson, T. (2013). A Climate for Academic Success: How School Climate Distinguishes Schools That Are Beating the Achievement Odds. Full Report. In *California Comprehensive Center at WestEd*. California Comprehensive Center at WestEd. <https://eric.ed.gov/?id=ED559741>
- Vygotsky, L. S. (1987). Thinking and speech. In R. W. Rieber & A. S. Carton (Eds.), *The collected works of L.S. Vygotsky, Volume 1: Problems of general psychology* ((Original work published 1934), pp. 39–285). Plenum Press.
- Walker, C. M. (2011). What's the DIF? Why Differential Item Functioning Analyses Are an Important Part of Instrument Development and Validation. *Journal of Psychoeducational Assessment*, 29(4), 364–376. <https://doi.org/10.1177/0734282911406666>
- Welner, K., & Farley, A. (2010). *Confronting systemic inequity in education: High impact strategies for philanthropy*. National Committee for Responsive Philanthropy.
- Worthington, R. L., & Whittaker, T. A. (2006). *Scale development research: A content analysis and recommendations for best practices*. *The Counseling Psychologist*.
- Wright, B. D. (1996). Comparing Rasch measurement and factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 3(1), 3–24. <https://doi.org/10.1080/10705519609540026>

- Wright, B. D., & Douglas, G. A. (1977). Best procedures for sample-free item analysis. *Applied Psychological Measurement, 1*(2), 281–295.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. MESA press.
- WXY Studio. (2018). *D15 Diversity Plan Final Report*. http://d15diversityplan.com/wp-content/uploads/2019/06/190620_D15DiversityPlan_FinalReport.pdf
- Ye, C., & Wang, Y. (2017). *ED School Climate Surveys (EDSCLS) Psychometric Benchmarking Technical Report* (p. 37). American Institutes for Research.
https://www.nassauboces.org/site/handlers/filedownload.ashx?moduleinstanceid=21724&dataid=32389&FileName=December%202017%20EDSCLS_Psychometric_Benchmarking_Technical_Report.pdf
- Youngs, C. S., & Youngs, G. A. (2001). Predictors of Mainstream Teachers' Attitudes Toward ESL Students. *TESOL Quarterly, 35*(1), 97–120. <https://doi.org/10.2307/3587861>
- Zieky, M. (1993). DIF statistics in test development. In P. W. Holland & H. Wainer, *Differential Item Functioning* (pp. 337–347). Erlbaum.
- Zlotnick, C., Goldblatt, H., Birenbaum-Carmeli, D., Dishon, Y., Taychaw, O., & Shadmi, E. (2019). The impact of adolescents' racial and ethnic self-identity on hope. *Health & Social Care in the Community, 27*(5), e705–e715. <https://doi.org/10.1111/hsc.12795>
- Zopluoglu, C., & Davenport Jr, E. (2017). A Note on Using Eigenvalues in Dimensionality Assessment. *Practical Assessment, Research, and Evaluation, 22*(1).
<https://doi.org/10.7275/e7gh-0785>
- Zullig, K. J., Collins, R., Ghani, N., Patton, J. M., Scott Huebner, E., & Ajamie, J. (2014). Psychometric support of the school climate measure in a large, diverse sample of adolescents: A replication and extension. *Journal of School Health, 84*(2), 82–90.

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

Zullig, K. J., Koopman, T. M., Patton, J. M., & Ubbes, V. A. (2010). School climate: Historical review, instrument development, and school assessment. *Journal of Psychoeducational Assessment, 28*(2), 139–152.

APPENDIX 1.

**Word cloud reflecting most frequent terms used in the definitions of sociocultural
integration and related constructs**



A word cloud of terms related to sociocultural integration. The words are arranged in a roughly rectangular shape, with 'culture' being the largest and most central. Other prominent words include 'language', 'multilingualism', 'value', 'power', and 'bias'. Smaller words like 'asset', 'discrimination', 'experiences', 'family', 'identity', 'justice', 'learning', 'opportunity', 'policies', 'representation', 'society', and 'teaching' are also present.

asset bias community context
culture discrimination experiences
family identity justice language
learning multilingualism opportunity
policies power representation society
teaching value

APPENDIX 2.

Facet and level combinations used for SCI-T item design

	SID L	SID M	SID H	PSE L	PSE M	PSE H	APE L	APE M	APE H
SID L	X	X	X	✓	X	X	✓	X	X
SID M	X	X	X	✓	✓	X	✓	✓	X
SID H	X	X	X	✓	✓	✓	✓	✓	✓
PSE L	X	✓	✓	X	X	X	✓	✓	X
PSE M	X	✓	✓	X	X	X	✓	✓	X
PSE H	X	X	✓	X	X	X	✓	✓	✓
APE L	✓	✓	✓	✓	✓	X	X	X	X
APE M	X	✓	✓	✓	✓	X	X	X	X
APE H	X	X	✓	✓	✓	✓	X	X	X

Note. SID stands for the first facet, “Self-Identification,” PSE for “Promotion of Self-Esteem,” and APE for “Agency for Power Equity.” The letters that follow each facet refer to the level; L stands for “Low,” M for “Moderate,” and H for “High” of each facet.

APPENDIX 3.

SCI-T Instrument Used in Cognitive Lab

Scenario	Facet/level combination	Scenario
SCIT-1.1	SCIT-1.1: HHH: SID_H + PSE_H + APE_H	Gloria acts uncomfortable when in the company of people from ethnic groups different than hers, for example, not looking people in the eye or being more quiet than usual. When discussing current events related to racial discrimination in the United States, Gloria tries to be politically neutral by not stating her opinion.
SCIT-1.2	SCIT-1.2 LLL: SID_L + PSE_L + APE_L	David avoids conversations about race and ethnicity because they are uncomfortable for him. In the classroom, he tracks the high-performing students and, depending on the cohort, this group may or may not be diverse in terms of ethnic and cultural background.
SCIT-1.3	SCIT-1.3: HHH: SID_H + PSE_H + APE_H	Laura shares with parents her experiences about being discriminated against and witnessing discrimination of others. For example, she may ask about their preferred language and then explain how standardized tests may be biased against their child because of their cultural or ethnic background.
SCIT-1.4	SCIT-1.4: LLL: SID_L + PSE_L + APE_L	Gloria is cordial to all her school peers, but it appears that she is only close to those who are the same ethnicity as her. She tells her friends how she worries all the time about saying things that might be considered prejudiced or biased.
SCIT-1.5	SCIT-1.5: HHH: SID_H + PSE_H + APE_H	Sebastian has proposed several times that school staff should get PD or other training on implicit bias and multicultural education. He is vocal about the staff representing the communities they serve and has asked about the diversity of the staff working at the school.
SCIT-2.1	SCIT-2.1: HHH: SID_H + PSE_H + APE_H	Laura sets time aside every day to reflect on whether she demonstrated to her students that she has high expectations for all of them. She also keeps track of misconducts to make sure that she did not punish students from an ethnic group disproportionately.

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

SCIT-2.2	SCIT-2.2 LLL: SID_L + PSE_L + APE_L	Gloria acts differently when around people whose ethnicity is different from hers, so she tries to fit in by toning down her personality to make others feel comfortable. Gloria gets anxious and experiences frustration when she has conferences with families whose culture is different from her own. She's never asked the parents if they feel nervous or frustrated when they meet with her.
SCIT-2.3	SCIT-2.3: HHH: SID_H + PSE_H + APE_H	When meeting a parent or guardian, Sebastian demonstrates his knowledge of their culture by commenting on relevant recent events or news. He is also known to provide or encourage parents to bring an interpreter to conferences when they are of a different ethnic group.
SCIT-2.4	SCIT-2.4: HHH: SID_H + PSE_H + APE_H	Laura encourages students to use the language they feel comfortable with and has strategies to help the class achieve shared understanding when multiple languages are participating. She purposely mixes students in groups that are ethnically diverse for classroom activities.
SCIT-2.5	SCIT-2.5: HHH: SID_H + PSE_H + APE_H	Laura provides multi-ethnic magazines, newspapers, and books for children to cut from or read at leisure (e.g., <i>Ebony</i> magazine, <i>Latino</i> magazine). She talks about injustice in society to her students and encourages conversation about current events related to racial discrimination in the United States.
SCIT-2.6	SCIT-2.6 LLL: SID_L + PSE_L + APE_L	David has not taken an Implicit Association Test. In the classroom, while trying to make clarifications for the culturally and ethnically diverse students in his classroom, he has inadvertently imitated their speech patterns.
SCIT-3.1	SCIT-3.1 LLL: SID_L + PSE_L + APE_L	When meeting a new student, parent, or peer whose ethnicity is different to his, David asks them where they are from. He's experienced more than one awkward moment with those parents because of misunderstandings in communication. He regrets not having sought counsel from supervisors about dealing with cultural misunderstandings with families.
SCIT-3.2	SCIT-3.2 LLL: SID_L + PSE_L + APE_L	If it's not easy to incorporate, David does not turn a lesson into an opportunity for students to share cultural differences in foods, dress, family life, or beliefs. David does not discuss issues like racial bias and cultural appropriation with students.

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

- SCIT-3.3 SCIT-3.3: HHH: SID_H + PSE_H + APE_H Sebastian does not shy away from talking about his ethnic and cultural background and his customs. He is also interested in the background of others. For example, he makes a list every school year of the ethnic groups represented by the students in his classroom, and he shares it with his students.
- SCIT-3.4 SCIT-3.4 LLL: SID_L + PSE_L + APE_L Gloria doesn't want to be identified as someone from a particular group. She rarely talks about her ethnic and cultural background. She has not spoken to her students about the surge in deportation of undocumented immigrants. She prefers to wait for the school leadership to tell her what she should do or say.
- SCIT-3.5 SCIT-3.5: HHH: SID_H + PSE_H + APE_H Laura speaks freely about racial bias and cultural appropriation with her students. For example, when an ethnic joke is spoken, Laura discusses its meaning with students and encourages them to determine if statements are appropriate and valid.
- SCIT-3.6 SCIT-3.6: HHH: SID_H + PSE_H + APE_H When meeting a parent or guardian, Sebastian demonstrates his knowledge of their culture by commenting on relevant recent events or news. He is also known to provide or encourage parents to bring an interpreter to conferences when they are of a different ethnic group.
-

APPENDIX 4.

Scenario Revisions After Cognitive Lab

Scenario	Average Rating (SD) ¹	Comments and feedback	Modification rationale
1.1 Gloria acts uncomfortable when in the company of people from ethnic groups different than hers, for example, not looking people in the eye or being more quiet than usual. When discussing current events related to racial discrimination in the United States, Gloria tries to be politically neutral by not stating her opinion.	1.45 (0.69)	<p><i>"We should all be politically neutral. We are at schools, we don't know what our parents think ... so no matter what, we should stay politically neutral."</i></p> <p><i>"It makes sense to me. To me, it's about this woman who can be involved in other cultures in a polite and a normal manner."</i></p> <p><i>"Letter a), I think the choice was slightly confusing. I act more so ... more what? I mean, I feel there was a slight confusion."</i></p>	<p>Support for Gloria's perspective suggests that the scenario is authentic, provides support.</p> <p>Response options are not clear, create hesitation in answering.</p>
1.2 David avoids conversations about race and ethnicity because they are uncomfortable for him. In the classroom, he tracks the high-performing students and, depending on the cohort, this group may or may not be diverse in terms of ethnic and cultural background.	3.00 (1.00)	<p><i>"I see it as two separate things, but then the more I read it and this is what I took from it: that he doesn't like looking at the racial and ethnic factors but then his group that he is divided may already be racially and ethnically divided."</i></p> <p><i>"Why would he track the high-performing students, why not the others? This is a little bit, what is the focus on this particular group?"</i></p> <p><i>"I thought it was easy because I enjoy talking (that might sound weird) about my race ethnicity, culture. I don't understand the tracking part, so I might or might not be. To me, I understood it, it didn't matter. It wasn't me either way."</i></p>	<p>Should I reverse this and say low-performing students?</p> <p>Support for the scenario.</p>

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

1.3 Laura shares with parents her experiences about being discriminated against and witnessing discrimination of others. For example, she may ask about their preferred language and then explain how standardized tests may be biased against their child because of their cultural or ethnic background.

2.00 (1.10)

"I didn't see two different questions on this one. This one felt like to me it was pretty cut and dry"

"I was growing up bilingual and it was really tough for me ... going to an English-only school. And that is something that I share with my families I think I understand ... where it was difficult for you to answer because maybe you didn't have that kind of experience ... but you still want your parents to be informed about the biases that are against their kid and are out there."

"I was debating whether to put b) that I act about the same or that I don't act like her because I wouldn't share experiences about me personally being discriminated against because I haven't faced that personally, but I would often engage in conversations about discrimination with parents."

"I ended up putting c) but I was between ... I ended up deciding that because I haven't personally faced – just reading it again."

Support for the scenario.

The wording of Laura being discriminated against, even if adding that she would speak about witnessing that of others, may keep participants from choosing the answer that is at their correct level. As, if someone hasn't been discriminated against, they can't enact a High.

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

<p>1.4 Gloria is cordial to all her school peers, but it appears that she is only close to those who are the same ethnicity as her. She tells her friends how she worries all the time about saying things that might be considered prejudiced or biased.</p>	<p>1.09 (0.30)</p>	<p><i>"I think with all these, I got a picture of the person in my head. I picture an old White lady that doesn't want to... but she is really nice, super nice, and very polite but doesn't really engage in real conversation with those outside her [window?]."</i></p> <p><i>"These are all ones and twos ... so it was pretty straightforward."</i></p>	<p>Support for the scenario.</p>
<p>1.5 Sebastian has proposed several times that school staff should get PD or other training on implicit bias and multicultural education. He is vocal about the staff representing the communities they serve and has asked about the diversity of the staff working at the school.</p>	<p>2.27 (1.42)</p>	<p><i>"To me there's multiple things here; he's proposed the staff PD, which is one issue, then he has also been very vocal about the staff representing the community, then he asked about the diversity of the staff. I didn't mark an answer because I didn't know which one to answer."</i></p> <p><i>"I felt I was like him, but I put a 5. He is really vocal about it, and proactive, and I'm not quite like that. I am all for the diversity training and multicultural awareness, but ... I feel like I am more passive about it."</i></p> <p><i>"One of the questions I asked when I was interviewing at my current school was how diverse is the staff, and how does the staff look like in relation to the school, so it's always been a very important thing for me. I can relate to that."</i></p>	<p>The facets don't go well together, causing confusion. Ponder whether these correspond to different levels (High and Moderate).</p> <p>This suggests the scenario is representing the correct level, High.</p>
<p>2.1 Laura sets time aside every day to reflect on whether she demonstrated to her students that she has high expectations for all of them. She also keeps track of misconducts to make sure that she did not punish students from an ethnic group disproportionately.</p>	<p>3.00 (0.93)</p>	<p><i>"Maybe it's possible to have more nuanced answers. I have 100% girls, and 98% Hispanic Maybe it's purely not ethnic, but cultural background by regions, districts of Mexico, rather than ethnic differences."</i></p>	<p>This is something I struggle with throughout, when to say ethnic or racial, to use them together, and not to leave aside the linguistic dimension.</p>

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

		<p><i>"Are there other differences, single family home, all boys/girls?"</i></p> <p><i>"Part of this is looking at philosophically with the lens of misconduct and punishment because you are already taking a stance [with the item wording]"</i></p>	<p>Bringing up punishment can be an unnecessary complication – it's a stance.</p>
<p>2.2 Gloria acts differently when around people whose ethnicity is different from hers, so she tries to fit in by toning down her personality to make others feel comfortable. Gloria gets anxious and experiences frustration when she has conferences with families whose culture is different from her own. She's never asked the parents if they feel nervous or frustrated when they meet with her.</p>	<p>3.44 (1.42)</p>	<p><i>"Although sometimes there could have been something in the conference that surprised me or that it didn't go as I would have thought it would in, and then in thinking about it later I realized that there was a miscommunication between what I was saying and how the parent experienced it."</i></p>	<p>Teachers are realizing issues as they try to answer the items. In this way items are working as an intervention themselves.</p>
		<p><i>"I was a really confused by 'I act more so than Gloria.' I'm like, more what? More anxious, more frustrated ... I act more differently? Is that what you mean? I ask parents if they feel nervous SOMETIMES but not ALWAYS, so I put that I don't act like Gloria."</i></p>	<p>Confusing answer options. In addition, frequency qualifiers help in separating levels of a facet. Their use should be limited.</p>
		<p><i>"I think for me as a Latina, toning down is a thing we learn to do at a young age, because we learn how to navigate White culture and a lot of administrators and even teachers a lot. So, we have to code switch a lot."</i></p>	<p>Proof of concept.</p>
		<p><i>"There's a contradiction there. OR you could also qualify that anxious as she's really good at code switching. In that case, I want to be more like her. I got confused about what toning down her personality, what that actually means"</i></p>	<p>Confusing use of toning down personality. Participants suggest that the problem may be the word "personality" and put forth other ways of saying the same thing, like <i>hide who you are</i>, or <i>tries to fit in to make others comfortable</i>.</p>

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

2.3 When meeting a parent or guardian, Sebastian demonstrates his knowledge of their culture by commenting on relevant recent events or news. He is also known to provide or encourage parents to bring an interpreter to conferences when they are of a different ethnic group. 2.89 (1.17)

"That last sentence, in our school we arrange for a translator. We make sure that it's someone who the parent is comfortable with."

"You can be a different ethnicity but still speak the same language"

"Something more meaningful would not be recent news or events. So, in my case, when I was in South Dakota, knowing just the students' backgrounds and the history of where those communities come from."

"If you are trying to measure cross-cultural competence, I think you are trying to say 'this is somebody who is trying to go out of his way to meet the needs' but at the same time it's like, well, if you are asking the parent to go find an interpreter then it's kind of not like what you thought."

Teachers expressed that they are not in charge of making sure there are interpreters/translators available for parents who need them. The last part of the scenario is unrealistic.

Ethnic group must be removed. Same discussion from other items in terms of use of race/ethnicity and linguistic diversity.

This signals that the first part of the scenario is a Moderate level rather than High (history>recent news>nothing). There seems to be a mismatch in the level of both parts of the scenario. Moderate and Low perhaps.

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

2.4 Laura encourages students to use the language they feel comfortable with and has strategies to help the class achieve shared understanding when multiple languages are participating. She purposely mixes students in groups that are ethnically diverse for classroom activities. 1.70 (1.34)

"Normally, that's what I do personally in my classroom when there's different cultures and different linguistics... I do it to learn diversity on different countries like... I think this is positive, in linguistic... to share and understand the language. I like the use of this strategy."

Proof of concept. Revise use of ethnically diverse, it should be linguistically in this scenario.

"...there are times when you want heterogeneous groups and then sometimes when you don't want to mix them... You have to be thinking about what the goal is for the teaching and what you take out of the experiences by mixing the groups ethnically..."

What is the purpose of the strategy at that point? Contextualization was a big problem with teachers deciding on their responses, they often asked for more context (in other groups) to help them choose a response.

"Some of these questions almost need something that begins by qualifying the context. 'Depending on the goal of teaching,' as opposed to these generalized statements that presume everything like this is what happens in all situations."

3.1 When meeting a new student, parent, or peer whose ethnicity is different to his, David asks them where they are from. He's experienced more than one awkward moment with those parents because of misunderstandings in communication. He regrets not having sought counsel from supervisors about dealing with cultural misunderstandings with families. N/A

"What I found in my school, I'm the minority. Everyone's culture is different than mine. So that wouldn't be my first response, because I don't assume that they are any different than the other kids in my classroom."
"A lot of immigration issues are at play at Arizona, so why is he asking that? Just based on the color? Based on the race that he thinks they are? I think I'm a little more sensitive to that coming from Arizona. And if you ask those types of questions, you may lose that family."
"In L.A., everyone is from different areas ... So sometimes if they've got a cool or interesting accent, I'll ask them where they are from or

I'm confused as to what to make of the first comment. On the one, I would expect teachers who are minority to answer differently, on the other, I'm not sure what White teachers interpreted here. The second comment, however, provides support for the idea that this is a Low level of the facet; you wouldn't ask this and in some regions/cities it might be even worse to do so. The third comment provides

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

		<p><i>something like that. I think it's more normal. And I'm the admissions counselor, so I get to know them and it's more like an ice breaker sometimes."</i></p>	<p>support for the regional perspective on variation of responses.</p>
		<p><i>"David is a real person." "People ask that all the time. Where are you from, where's your family from? I'm from the United States. I feel people ask that all the time based off appearances."</i></p>	<p>Support for the scenario.</p>
		<p><i>"All I know is that he had an awkward encounter with the family. So, why did he have that awkward encounter? Does he not get to know the families? Is this his first time meeting them?"</i></p>	<p>The wording "awkward moment" is not descriptive enough, perhaps misleading.</p>
		<p><i>"After reading it a couple of times I understood." "What I got from it, putting myself in David's shoes, would I act like David or would I act differently from David?"</i></p>	<p>The instructions that accompany the items should be modified using the expressions teachers used to rationalize their responses. Perhaps go as far as asking participants to read scenarios twice (or more).</p>
<p>3.2 If it's not easy to incorporate, David does not turn a lesson into an opportunity for students to share cultural differences in foods, dress, family life, or beliefs. David does not discuss issues like racial bias and cultural appropriation with students.</p>	<p>2.40 (1.71)</p>	<p><i>"Maybe but it would be harder to steer away from talking about culture when you have a mix of two [cultures]." "It also depends on grade level. At middle school or maybe at elementary you can start talking about racial bias and stuff."</i></p>	<p>This teacher suggests that in DLS you have a mix of cultures that requires/inspires the topic to be discussed.</p>

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

		<p><i>"I think it happens in a lot of schools, but I think it's kind of changing a little. People are using those opportunities more to talk about culture."</i></p>	<p>Support for the scenario.</p>
<p>3.3 Sebastian does not shy away from talking about his ethnic and cultural background and his customs. He is also interested in the background of others. For example, he makes a list every school year of the ethnic groups represented by the students in his classroom, and he shares it with his students.</p>	<p>2.90 (0.88)</p>	<p><i>"I was kind of swayed by the example a little bit. I was like 'well, I don't shy away from it, but I don't do that.' So, then I was 'well, do I actually act like Sebastian?' But I ended up kind of ignoring the example and focusing on the first part."</i></p> <p><i>"The example is a little extreme."</i></p>	<p>The example seems too extreme for the facet. It is meant to represent the unattainable High level, but seems a little odd in reality.</p>
<p>3.4 Gloria doesn't want to be identified as someone from a particular group. She rarely talks about her ethnic and cultural background. She has not spoken to her students about the surge in deportation of undocumented immigrants. She prefers to wait for the school leadership to tell her what she should do or say.</p>	<p>2.44 (1.33)</p>	<p><i>"I feel like it has two separate things. But I work with the youngest group of children in the school so it's a topic that hasn't come up and thankfully none of my students have gone through that. But if I had the opportunity, I would probably talk about it, just like you said adapted to their age, and I'm sure they'll get it Even though they are very young they know that someone's mommy is going away for a long time or they might have to go too."</i></p> <p><i>"I was kind of feeling like Gloria was having the fear of deportation, not that she didn't want to talk about." "That's how I read it too."</i></p>	<p>Support for the scenario.</p> <p>There's a misunderstanding about whether Gloria fears to be deported or just fears talking about this occurring in the community.</p>

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

3.5 Laura speaks freely about racial bias and cultural appropriation with her students. For example, when an ethnic joke is spoken, Laura discusses its meaning with students and encourages them to determine if statements are appropriate and valid.

2.13 (1.13)

"I would even go a little more and talk about with the students why these types of jokes in general are not appropriate and why are we making fun of somebody's ethnic background. Just try to encourage them to step away from any kind of ethnic jokes at all ... and do different little dramas or something like that about what about when you hear that about your own ethnic background how does it make you feel, so why are you doing that to somebody else?"

Additional material for other examples.

"You don't want to put as much influence in what one kid says, but then what is the other thing that you're doing over here. It's reactive but then it's the proactive too."

This suggests that the scenario might not actually be representing the highest level of the facet.

"I don't know how to feel about this one The first part, I understand, and I get, and I like, but I don't know about the appropriate[ness] and validity of it. That's the scary part. Even if they are just middle school kids, you are giving them the power to say that it's valid to say."

Absolutely remove the words valid and appropriate. I mention in the audio that it was an error I made when elaborating this scenario.

3.6 When meeting a parent or guardian, Sebastian demonstrates his knowledge of their culture by commenting on relevant recent events or news. He is also known to provide or encourage parents to bring an interpreter to conferences when they are of a different ethnic group.

2.63 (1.51)

"And the second part with the conferences, just because they are from a different ethnic background doesn't necessarily mean that they need an interpreter; that's where I got hung up, so I'm not quite sure how to answer, to be honest."

I mention in the audio that it was an error I made when elaborating this scenario.

"I always let my principal know that I'm meeting, especially if it's not during parent teacher conference time, if it's an outside meeting we let her know and she'll say or I'll ask, 'would you like to be a part of this meeting to help them?'"

Other ways to rephrase.

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

"In the teacher parent conference, you don't want just anybody discussing records with anybody else. Our front office staff are both bilingual, so if they need a translator, we'll pull one of them in."

Teachers can't just bring someone in to serve as interpreter.

¹ Statistics were calculate using for n≤11 for each of the three groups used in the cognitive labs. Statistics are not available for items that were not administered due to time constraints.

APPENDIX 5.

SCI-T Pilot Facet and Level Combinations and Scores

Scenario	Level by facet	Facet/level combinations	Score
SCIT-1	High-High-High	SID_H1 + PSE_H1 + APE_H3	3 + 3 + 3 = 9
SCIT-2	High-High-High	SID_H2 + PSE_H2 + APE_H2	3 + 3 + 3 = 9
SCIT-3	High-Moderate-High	SID_H3 + PSE_M1 + APE_H1	3 + 2 + 3 = 8
SCIT-4	High-Moderate-High	SID_H2 + PSE_M2 + APE_H2	3 + 2 + 3 = 8
SCIT-5	High-High-Moderate	SID_H1 + PSE_H3 + APE_M1	3 + 3 + 2 = 8
SCIT-6	High-High-Moderate	SID_H3 + PSE_H2 + APE_M2	3 + 3 + 2 = 8
SCIT-7	Moderate-Moderate-Moderate	SID_M1 + PSE_M1 + APE_M1	2 + 2 + 2 = 6
SCIT-8	Moderate-Moderate-Moderate	SID_M2 + PSE_M3 + APE_M2	2 + 2 + 2 = 6
SCIT-9	Moderate-Low-Moderate	SID_M3 + PSE_L1 + APE_M1	2 + 1 + 2 = 5
SCIT-10	Moderate-Low-Moderate	SID_M2 + PSE_L2 + APE_M3	2 + 1 + 2 = 5
SCIT-11	Moderate-Moderate-Low	SID_M1 + PSE_M1 + APE_L1	2 + 2 + 1 = 5
SCIT-12	Moderate-Moderate-Low	SID_M3 + PSE_M2 + APE_L2	2 + 2 + 1 = 5
SCIT-13	Low-Low-Low	LLL: SID_L1 + PSE_L1 + APE_L1	1 + 1 + 1 = 3
SCIT-14	Low-Low-Low	LLL: SID_L3 + PSE_L2 + APE_L2	1 + 1 + 1 = 3

APPENDIX 6.

SCI-T Instrument Used in Pilot

Scenario	Facet/Level Combination	Scenario
SCIT-1	HHH: SID_H1 + PSE_H1 + APE_H3	Teacher X is aware and distressed about how systemic, unearned racial/linguistic advantage and conferred dominance plays out at school and the community. She actively seeks ways to integrate lessons and materials about current issues affecting Latino people and communities into the curriculum. Teacher X's students consider her classroom a safe environment where they can be heard. Students feel comfortable using the board to draw things they are trying to describe, or switch languages at any time they need.
SCIT-7	MMM: SID_M1 + PSE_M1 + APE M1	Teacher X perceives racial/linguistic advantage and conferred dominance in some situations at school and the community. She does not consider them systemic. Teacher X does not include in her lesson plans items that encourage pride in her students' cultures, but she makes connections in class that serve that purpose. Teacher X has discussed with peers but not parents how immigrant students might do worse in standardized tests because of their linguistic or cultural/ethnic background.
SCIT-2	HHH: SID_H2 + PSE_H2 + APE_H2	In the presence of representations of different cultures, Teacher X has spoken to his students about the difference between cultural appreciation and the unacknowledged or inappropriate adoption of cultural features. When Teacher X meets a new student or parent, he asks them which language they prefer to engage in conversation. He lets them know the languages he speaks and the interpretation services available at the school. After last year's events, Teacher X has encouraged immigrant students in his class to discuss their feelings and fears around ICE raids.

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

SCIT-12	MML: SID_M3 + PSE_M2 + APE L2	If a joke charged with bias is spoken in his classroom, Teacher X is uncomfortable walking his students through its negative assumptions. When Teacher X meets a new student or parent with Spanish surname, he immediately addresses them with "Hola, ¿cómo estás?" After last year's events, Teacher X defers to the school's principal to discuss anything related to ICE raids. He will not engage in a conversation about the topic with students or parents.
SCIT-13	LLL: SID_L1 + PSE_L1 + APE L1	Teacher X does not conceive that racial/or linguistic dominance could affect him/her or his students. He regrets not having sought counsel from peers or supervisors when dealing with linguistic or cultural misunderstandings with students or families. Teacher X defers the responsibility of encouraging pride in students' cultures to the school.
SCIT-3	HMH: SID_H3 + PSE_M1 + APE H1	If a joke charged with bias is spoken in her classroom, Teacher X immediately walks her students through its negative assumptions. Teacher X does not include in her lesson plans items that encourage pride in her students' cultures, but she makes connections in class that serve that purpose. Teacher X has explained to parents how standardized tests may be biased against immigrant students because of their linguistic or cultural/ethnic background.
SCIT-10	MLM: SID_M2 + PSE_L2 + APE M3	Teacher X wants to advocate for equity in her classroom, but she is just one teacher with a lot of students in her class. Diversifying lessons is difficult for him/her. In the presence of representations of different cultures, Teacher X feels uncomfortable about cultural features being adopted inappropriately. When Teacher X meets a new student or parent who has brown skin, she immediately addresses them with "Hola, ¿cómo estás?"
SCIT-4	HMH: SID_H2 + PSE_M2 + APE H2	In the presence of representations of different cultures, Teacher X has spoken to his students about the difference between cultural appreciation and the unacknowledged or inappropriate adoption of cultural features. After last year's events, Teacher X has encouraged immigrant students in his class to discuss their feelings and fears around ICE raids. When he meets a new student or parent with Spanish surname, he immediately addresses them with "Hola, ¿cómo estás?"

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

SCIT-5	HHM: SID_H1 + PSE_H3 + APE M1	Teacher X is aware and distressed about how systemic, unearned racial/linguistic advantage and conferred dominance plays out at school and the community. When she sees students' languages being devalued in school, Teacher X speaks up no matter who the offender is. Teacher X has discussed with peers but not parents how immigrant students might do worse in standardized tests because of their linguistic or cultural/ethnic background.
SCIT-9	MLM: SID_M3 + PSE_L1 + APE M1	If a joke charged with bias is spoken in his classroom, Teacher X is uncomfortable walking his students through its negative assumptions. Teacher X has discussed with peers but not parents how immigrant students might do worse in standardized tests because of their linguistic or cultural/ethnic background. Teacher X defers the responsibility of encouraging pride in students' cultures to the school.
SCIT-11	MML: SID_M1 + PSE_M1 + APE L1	Teacher X does not include in her lesson plans items that encourage pride in her students' cultures, but she makes connections in class that serve that purpose. Teacher X regrets not having sought counsel from peers or supervisors when dealing with linguistic or cultural misunderstandings with students or families. Teacher X perceives racial/linguistic advantage and conferred dominance in some situations at school and the community. She does not consider them systemic.
SCIT-14	LLL: SID_L3 + PSE_L2 + APE L2	When Teacher X meets a new student or parent who has brown skin, he immediately addresses them with "Hola, ¿cómo estás?" After last year's events, Teacher X defers to the school's principal to discuss anything related to ICE raids. He will not engage in a conversation about the topic with students or parents. If a joke charged with bias is spoken in his classroom, Teacher X will change the topic.
SCIT-6	HHM: SID_H3 + PSE_H2 + APE M2	When Teacher X meets a new student or parent, she asks them which language they prefer to engage in conversation. If a joke charged with bias is spoken in her classroom, Teacher X immediately walks her students through its negative assumptions. She lets them know the languages she speaks and the interpretation services available at the school. After last year's events, Teacher X has asked fellow teachers whether she should talk to students about ICE raids.

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

SCIT-8

MMM: SID_M2 + PSE_M3 + APE M2

In the presence of representations of different cultures, Teacher X feels uncomfortable about cultural features being adopted inappropriately. Teacher X is worried and upset when she sees students' languages being devalued in school, but she does not speak up. After last year's events, Teacher X has asked fellow teachers whether she should talk to students about ICE raids.

APPENDIX 7.

SCI-T Instrument Used in Main Study

Scenario	Facet/level combination	Scenario
SCIT-1	HHH: SID_H1 + PSE_H1 + APE_H4	The teacher is aware and distressed about how systemic, unearned racial advantage plays out at their school and community. The teacher is committed to an inclusive curriculum that integrates cultural and ethnic diversity all year long. They actively seek ways to integrate lessons and materials about their underrepresented students' backgrounds into the curriculum. The teacher talks about issues of racial justice, topics like Black voices, stories, and achievements regularly. They are talking to students about activism, the Black Lives Matter movement, and racial injustice.
SCIT-7	MMM: SID_M1 + PSE_M1 + APE_M5	The teacher perceives racial advantage in some school and community situations. They do not consider these advantages to be systemic. The teacher includes a couple of lessons in the curriculum meant to encourage pride in some of the cultures represented by their students. The teacher was concerned about students who missed class, reported them according to protocol, and subsequently reached out to students and their caretakers to ask what may have caused the absence.
SCIT-2	HHH: SID_H2 + PSE_H2 + APE_H2	The teacher uses a Facebook post showing a model wearing a traditional Native American headdress to speak about the difference between cultural appreciation and cultural appropriation (the unacknowledged or inappropriate adoption of cultural features). The teacher asks students to question whether the portrayal is insulting or insensitive and think about who might be in a qualified position to say whether the portrayal is or is not offensive. When meeting a new student or parent, the teacher asks if they speak multiple languages. If they do, the teacher asks which language they prefer for conversation, with the goal of accommodating that preference, if possible. The teacher has encouraged immigrant students, and students with immigrant family members or caretakers, to share with the teacher privately or to confide in fellow classmates any feelings and fears around ICE raids.

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

SCIT-12	MML: SID_M3 + PSE_M2 + APE_L2	The teacher holds pride in their own racial/ethnic group. They prefer associating with and promoting opportunities to others who identify with this group. The teacher speaks no language other than English. When meeting a new student or parent with a Spanish surname, the teacher instinctively addresses them with "Hola, ¿cómo estás?" The teacher defers to the school's principal to discuss anything related to ICE raids, finding it inappropriate to engage in conversation about this topic with students or parents.
SCIT-13	LLL: SID_L1 + PSE_L3 + APE_L4	The teacher does not believe racial advantage could affect them or their students. A student laughed at a classmate for switching to their native language to explain an idea. The teacher didn't say anything at the time, not knowing how to address it. Later in the day, the teacher checked in with the affected student. When students have brought up the Black Lives Matter movement, the teacher steered the conversation, saying that "all lives matter" and requesting that students get back on task.
SCIT-3	HMH: SID_H3 + PSE_M1 + APE_H3	The teacher feels positively about the racial/ethnic groups they identify with. That positive attitude extends to the groups that others identify with, as well as being committed to antiracist practices. The teacher includes a couple of lessons in the curriculum meant to encourage pride in some of the cultures represented by their students. The teacher consistently works to meet the needs of every student in the classroom and integrates systems that allow for reflection and refinement. The teacher consistently amplifies student voice by differentiating lessons for engagement, representation, and expression.
SCIT-10	MLM: SID_M2 + PSE_L1 + APE_M6	The teacher advised students not to wear other cultures' traditional costumes for Halloween. When asked by a student, the teacher was unable to explain why it is inappropriate to use heritage elements from any non-dominant cultural group as a costume. The teacher defers to school administrators the responsibility of encouraging pride in students' cultures. The teacher had a few misunderstandings with non-native-English-speaking parents and regrets not having sought counsel from peers when dealing with parents who speak languages the teacher does not. Now the teacher tries to have someone available to translate such conversations, should translation be needed.
SCIT-4	HMH: SID_H3 + PSE_M2 + APE_H7	The teacher feels positively about the racial/ethnic groups they identify with. That positive attitude extends to the groups that others identify with, as well as being committed to antiracist practices. The teacher speaks no language other than English. When meeting a new student or parent with a Spanish surname, the teacher instinctively addresses them with "Hola, ¿cómo estás?" The teacher is involved in a lot of work in equity in education. Reaching out to caregivers and other educators, the teacher has created an antiracist discussion group for their county. The teacher engages in readings and develops monthly action plans with concrete antiracist actions.

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

SCIT-5	HHM: SID_H1 + PSE_H3 + APE_M3	The teacher is aware and distressed about how systemic, unearned racial advantage plays out at their school and community. A student laughed at a classmate for switching to their native language to explain an idea. The teacher reminded all students that classroom norms, established early in the school year, include respect for all diversities and all voices. The teacher then established a restorative space for the two students to talk and checked in with both students at the end of the day. The teacher advocates for equity in their classroom, to a certain extent. The teacher differentiates lessons based on students' special education needs. The teacher is not comfortable exerting more effort than the administration in pursuing equity.
SCIT-9	MLM: SID_M3 + PSE_L1 + APE_M4	The teacher holds pride in their own racial/ethnic group. They prefer associating with and promoting opportunities to others who identify with this group. The teacher defers to school administrators the responsibility of encouraging pride in students' cultures. When students brought up the topic, the teacher has engaged in conversation about the Black Lives Matter movement. The teacher did not know how to facilitate or enable students to lead those discussions.
SCIT-11	MML: SID_M1 + PSE_M1 + APE_L6	The teacher perceives racial advantage in some school and community situations. They do not consider these advantages to be systemic. The teacher includes a couple of lessons in the curriculum meant to encourage pride in some of the cultures represented by their students. The teacher had a few misunderstandings with non-native-English-speaking students and their parents. They believe parents should look into city or community resources to learn English. For the time being, the teacher avoids any non-urgent conversations with those parents.
SCIT-14	LLL: SID_L3 + PSE_L2 + APE_L7	Racial/ethnic affiliation is not something the teacher gives much thought to. The teacher refuses to engage in conversation on the topic. The teacher speaks no language other than English. When meeting a new student or parent who looks LatinX, the teacher addresses them with "Hola, ¿cómo estás? Is English OK?" The teacher is aware that racial injustice is something everyone needs to pitch in to solve. The teacher prefers to leave that work to others who are better equipped and have more time because they are overwhelmed by what is already on their plate.

MEASURING TEACHERS' PROMOTION OF SOCIOCULTURAL INTEGRATION

- SCIT-6 HHM: SID_H3 + PSE_H2 + APE_M2 The teacher feels positively about the racial/ethnic groups they identify with. That positive attitude extends to the groups that others identify with, as well as being committed to antiracist practices. When meeting a new student or parent, the teacher asks if they speak multiple languages. If they do, the teacher asks which language they prefer for conversation, with the goal of accommodating that preference, if possible. After last year's immigration challenges, the teacher asked colleagues if it's appropriate to talk to students about ICE raids.
- SCIT-8 MMM: SID_M2 + PSE_M3 + APE_M7 The teacher advised students not to wear other cultures' traditional costumes for Halloween. When asked by a student, the teacher was unable to explain why it is inappropriate to use heritage elements from any non-dominant cultural group as a costume. A student laughed at a classmate for switching to their native language to explain an idea. The teacher responded, "That's inappropriate," offering no explanation. The teacher checked in with the affected student later in the day. The teacher is trying to learn how to be a better ally by engaging in readings related to antiracism. The teacher is insecure about the actions they could carry out in their school and community but is looking at organizations that work on that mission. The teacher finds antiracism work hard but is committed to trying.
-

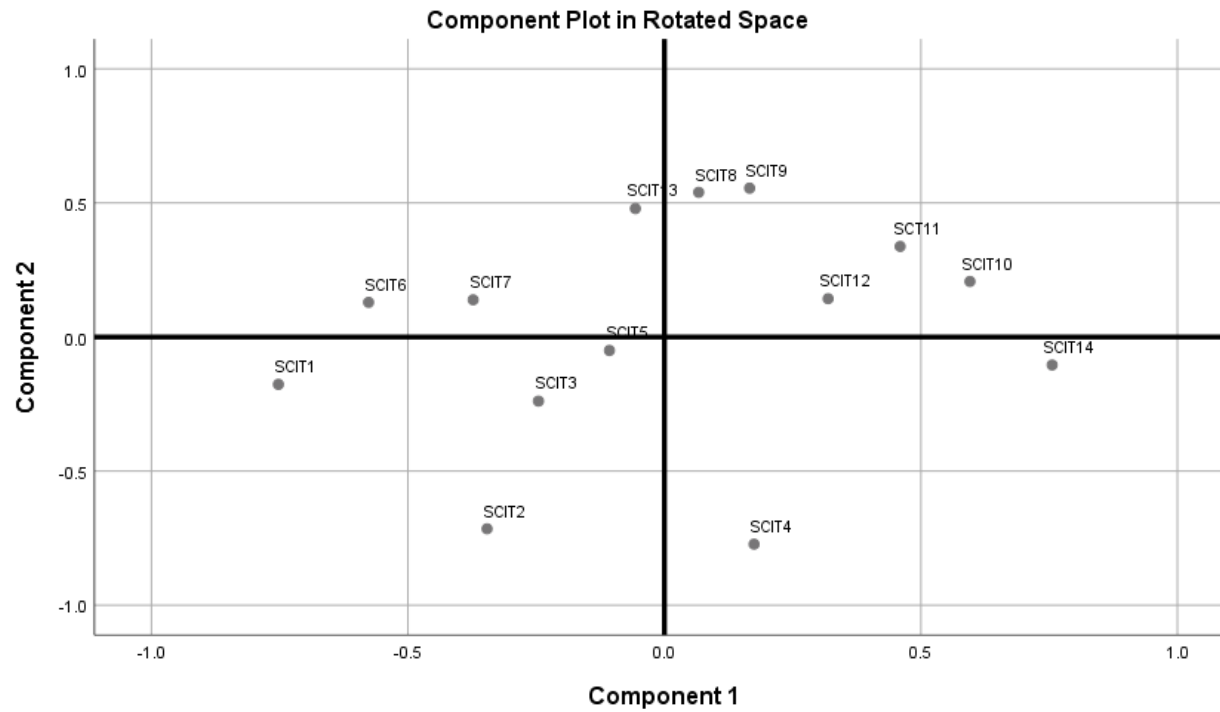
APPENDIX. 8

Pilot Study: Respondent Fit Statistics

Respondent ID	Measure	INFIT		OUTFIT	
		MNSQ	ZSTD	MNSQ	ZSTD
S1001	4.67	4.03	3.68	5.92	3.81
S2007	1.03	3.55	3.76	3.55	3.8
E1077	-1.21	3.55	3.5	3.51	3.43
E1033	0.7	3.01	3.55	3	3.57
E1008	1.01	2.87	3.41	2.9	3.47
E1071	-1.16	2.69	3.36	2.79	3.42

APPENDIX. 9

Pilot Study: Residual Analysis – Varimax Rotation Component Plot



APPENDIX. 10

Main Study: Residual Analysis – Varimax Rotation Component Plot

