

# Measuring the complexity of teachers' enactment of practice for equity: A Rasch model and facet theory-based approach

Author: Wen-Chia Claire Chang

Persistent link: <http://hdl.handle.net/2345/bc-ir:107345>

This work is posted on [eScholarship@BC](#),  
Boston College University Libraries.

---

Boston College Electronic Thesis or Dissertation, 2017

Copyright is held by the author, with all rights reserved, unless otherwise noted.

**BOSTON COLLEGE**  
Lynch School of Education

Department of  
Educational Research, Measurement, and Evaluation

MEASURING THE COMPLEXITY OF TEACHERS' ENACTMENT OF PRACTICE  
FOR EQUITY: A RASCH MODEL AND FACET THEORY-BASED APPROACH

Dissertation

by

WEN-CHIA CLAIRE CHANG

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

May 2017

© Copyright by Wen-Chia Claire Chang

2017

## ABSTRACT

### MEASURING THE COMPLEXITY OF TEACHERS' ENACTMENT OF PRACTICE FOR EQUITY: A RASCH MODEL AND FACET THEORY-BASED APPROACH

Wen-Chia C. Chang, Author

Larry H. Ludlow, Chair

Preparing and supporting teachers to enact teaching practice that responds to diversity, challenges educational inequities, and promotes social justice is a pressing yet daunting and complex task. More research is needed to understand how and to what extent teacher education programs prepare and support teacher candidates to enhance the achievement of all learners while challenging systematic inequity (Cochran-Smith, Ell, Ludlow, Grudnoff, & Aitken, 2014). One piece of empirical evidence needed is a measure that captures the extent to which teachers enact teaching practice for equity.

This study developed an instrument – the Teaching Equity Enactment Scenario Scale (TEES) - to measure the extent of equity-centered teaching practice by applying Rasch measurement theory (Rasch, 1960) and Guttman's facet theory (Borg & Shye, 1995). The research question addressed whether the TEES scale can measure teachers' self-reported enactment of practice for equity in a reliable, valid, and authentic manner. This study employed a three-phase design, comprising an extensive process of item development, a pilot study and a final full-scale administration. Fifteen scenario-style items were developed to capture the enactment levels of six interconnected principles of teaching practice for equity. Using the Rasch rating scale model the outcome was a 15-item TEES scale that reliably and validly measures increasing levels of teaching practice for equity progressing through low, moderate, and high levels of enactment. The distribution of the scenarios confirmed their hypothesized order and the

instrument development principles of Rasch measurement - unidimensionality, variation and a hierarchical order of the items, as well as a uniform continuum defining the construct. The scale also provides meaningful interpretations of what a raw score means regarding one's equity-centered teaching practice.

The overall findings suggest that the novel approach of combining Rasch measurement and facet theory can be successful in developing a scenario-style scale that measures a complex construct. Moreover, the scale can provide the evidence needed in research on preparing and supporting teachers to teach with a commitment to equity and social justice.

## ACKNOWLEDGMENTS

While nearing the end of this journey of pursuing a doctoral degree, I am overwhelmed by all the help, coaching, guidance, and encouragement from so many individuals whom I have had the privilege to work with, to learn from, and to share life. Words can never adequately capture my deep appreciation.

I want to first express my gratitude to my wonderful committee chair, Dr. Larry Ludlow. Thank you for advocating for me when I first transferred to the department. You have shown me what it is like to be a good mentor who cares for and wants the best for students. I also want to express my gratitude to my other committee members – Drs. Marilyn Cochran-Smith, Lexie Grudnoff, and Lauren Saenz. The dissertation journey was enjoyable, meaningful, and inspiring because of you. Your dedication, integrity, encouragement, kindness, humor, and insight inspired me to produce the kind of work that can be both personally meaningful and make a contribution to the field of education. I want to thank you for your feedback, which made me reflect, rethink, and work through the topics where I felt the least comfortable. I want to thank you for challenging and believing in me.

I additionally want to express my appreciation to other members of the Rethinking Initial Teacher Education (RITE) research team – Drs. Fiona Ell, Mary Hill, and Mavis Haigh. Your encouraging words, patience, feedback, and help made this dissertation study possible. I also want to thank my colleagues from the Educational Research, Measurement, and Evaluation and Curriculum and Instruction departments at Boston College. Your support and suggestions during the many difficult times of the dissertation mean the world to me. I want to thank my former supervisor, Angela Chen,

and the coworkers with whom I worked closely at Education Development Center (EDC). The opportunities, experience, and coaching that I received over the course of four years at EDC have been invaluable to this journey and to my career.

I can never express enough gratitude to my parents - thank you for letting me go and providing me the resources and freedom to explore and to cultivate my own path; thank you for being there when I was exhausted and disheartened; and, finally -- thank you for letting me being myself. To my friends, and especially friends from Masacote Dance Company, Jose Mateo Ballet Theater, and Jeannette Neill Dance Studio, I could never have gone so far without your friendship, support, understanding, and of course dance and movement. To those friends I have lost along the way of this doctoral program, I thank you for teaching me about life.

To my study participants and informants, I would like to thank you for sharing your suggestions, your teaching philosophy, and your passion about teaching. You inspire me and help to remind me of the ultimate reason why I do what I do.

## TABLE OF CONTENTS

<b>ABSTRACT</b> .....	<b>i</b>
<b>ACKNOWLEDGMENTS</b> .....	<b>iii</b>
<b>LIST OF TABLES</b> .....	<b>viii</b>
<b>LIST OF FIGURES</b> .....	<b>ix</b>
<b>CHAPTER ONE: INTRODUCTION</b> .....	<b>1</b>
<b>The Problem</b> .....	<b>1</b>
<b>Purpose of the Proposed Study</b> .....	<b>5</b>
<b>Significance of the Study</b> .....	<b>7</b>
<b>Summary</b> .....	<b>10</b>
<b>CHAPTER TWO: LITERATURE REVIEW</b> .....	<b>13</b>
<b>Teaching that Promotes Equity and Social Justice</b> .....	<b>14</b>
Principles of teachers' enactment of practice for equity.....	16
<b>Existing Instruments Measuring Aspects of Teaching for Equity and Social Justice</b> .....	<b>21</b>
<b>Review of Measurement Theories</b> .....	<b>30</b>
Classical Test Theory (CTT).....	30
Rasch measurement theory.....	40
Facet theory (FT) and the sentence mapping technique.....	57
<b>Putting it together: Combining Rasch measurement theory and facet theory</b> .....	<b>59</b>
<b>Summary</b> .....	<b>64</b>
<b>CHAPTER THREE: METHODOLOGY</b> .....	<b>65</b>
<b>Research Design</b> .....	<b>65</b>
<b>Participants</b> .....	<b>66</b>
<b>Sampling</b> .....	<b>66</b>
Pilot study.....	67
Final full-scale administration.....	68
<b>Procedure</b> .....	<b>69</b>
Step one: Reviewing the syntheses and clarifying the construct.....	69
Step two: Applying a systematic approach to developing rich narratives and descriptions on three levels of enactment for each facet.....	72
Step three: Determining the scale's structure for scenario development.....	74
Step four: Formulating sentence mapping structures and constructing scenarios.....	77
Step five: Conducting a pre-pilot and feedback sessions.....	81
<b>Data Collection</b> .....	<b>85</b>
Instrumentation.....	85
<b>Data Analysis</b> .....	<b>87</b>
<b>CHAPTER FOUR: RESULTS</b> .....	<b>92</b>
<b>Pilot Study</b> .....	<b>92</b>
Overview of participants' responses.....	92
Descriptive statistics.....	97
Rasch analyses.....	99
Item revision.....	113
<b>Final Full-Scale Administration</b> .....	<b>120</b>
Overview of participants' responses.....	120
Descriptive statistics.....	123



Rasch analyses .....	124
<b>Summary of Results .....</b>	<b>155</b>
<b>CHAPTER FIVE: DISCUSSION.....</b>	<b>156</b>
<b>Overview of Findings .....</b>	<b>156</b>
<b>Discussion of Findings.....</b>	<b>157</b>
<b>Limitations .....</b>	<b>164</b>
<b>Recommendations for Future Research.....</b>	<b>167</b>
<b>Implications and Conclusion .....</b>	<b>169</b>
<b>References.....</b>	<b>173</b>
<b>Appendix A: Construct Mapping Sheet.....</b>	<b>184</b>
<b>Appendix B: Collection of Practice for Six Facets.....</b>	<b>198</b>
<i>Facet One .....</i>	<i>198</i>
<i>Facet Two .....</i>	<i>203</i>
<i>Facet Three.....</i>	<i>207</i>
<i>Facet Four .....</i>	<i>213</i>
<i>Facet Five.....</i>	<i>217</i>
<i>Facet Six .....</i>	<i>220</i>
<b>Appendix C: General Descriptions for Six Facets (Version One) .....</b>	<b>224</b>
<i>Facet One .....</i>	<i>224</i>
<i>Facet Two .....</i>	<i>226</i>
<i>Facet Three.....</i>	<i>227</i>
<i>Facet Four .....</i>	<i>229</i>
<i>Facet Five.....</i>	<i>230</i>
<i>Facet Six .....</i>	<i>231</i>
<b>Appendix D: General Descriptions for Six Facets (Version Two).....</b>	<b>232</b>
<i>Facet One .....</i>	<i>232</i>
<i>Facet Two .....</i>	<i>233</i>
<i>Facet Three.....</i>	<i>234</i>
<i>Facet Four .....</i>	<i>235</i>
<i>Facet Five.....</i>	<i>236</i>
<i>Facet Six .....</i>	<i>237</i>
<b>Appendix E: Descriptions of Levels of Enactment for Six Facets .....</b>	<b>238</b>
<i>Facet One .....</i>	<i>238</i>
<i>Facet Two .....</i>	<i>241</i>
<i>Facet Three.....</i>	<i>244</i>
<i>Facet Four .....</i>	<i>247</i>
<i>Facet Five.....</i>	<i>249</i>
<i>Facet Six .....</i>	<i>252</i>
<b>Appendix F: Sentence Mapping Structure .....</b>	<b>255</b>
<i>Sentence mapping structure and three levels of scenarios: Facets one, two, and six.....</i>	<i>255</i>
<i>Sentence mapping structure and three levels of scenarios: Facets two, three, and six .....</i>	<i>257</i>
<i>Sentence mapping structure and three levels of scenarios: Facets three, four, and six .....</i>	<i>259</i>
<i>Sentence mapping structure and three levels of scenarios: Facets four, five, and six.....</i>	<i>261</i>
<i>Sentence mapping structure and three levels of scenarios: Facets one, five, and six.....</i>	<i>263</i>
<b>Appendix G: Guiding Questions for Feedback Sessions.....</b>	<b>265</b>

<b>Appendix H: Comparison of the Original and the Revised Scenarios.....</b>	<b>267</b>
<b>Appendix I: Pilot Survey .....</b>	<b>277</b>
<b>Appendix J: Final Survey.....</b>	<b>292</b>
<b>Appendix K: WINSTEPS Program Scripts .....</b>	<b>309</b>
<i>Pilot Study .....</i>	<b>309</b>
<i>Final Study (57 cases) .....</i>	<b>311</b>
<i>Final Study (52 cases) .....</i>	<b>313</b>
<b>Appendix L: Revised Teaching Equity Enactment Scenarios in “Difficulty” Order .....</b>	<b>315</b>
<b>Appendix M: Scoring Conversing Table .....</b>	<b>319</b>

## LIST OF TABLES

- Table 2.1 *Linkages between the Five Syntheses and the Six Principles of Practice*
- Table 3.1 *Scenario Development Structure*
- Table 3.2 *Survey Section Plan*
- Table 4.1 *Number Missing: Pilot Study*
- Table 4.2 *Number Missing by Country: Pilot Study*
- Table 4.3 *Number Missing by Response Categories of the Practice Item: Pilot Study*
- Table 4.4 *Descriptive Statistics for the Scenarios: Pilot Study*
- Table 4.5 *TEES Scenario Difficulty Estimates by Subgroups*
- Table 4.6 *Item Misfit: Pilot Study*
- Table 4.7 *Item Revisions for the Final Study*
- Table 4.8 *Revised Sentence Mapping Levels: Changes from Pilot to Final Survey*
- Table 4.9 *Number Missing: Final Study*
- Table 4.10 *Number Missing by Sample Sources: Final Study*
- Table 4.11 *Number Missing by Response Categories of the Practice Item: Final Study*
- Table 4.12 *Descriptive Statistics for the Scenarios: Final Study*
- Table 4.13 *Item Misfit: Final Study*
- Table 4.14 *Person Misfit: Final Study*
- Table 4.15 *Teaching Equity Enactment Scenario Scale Score Interpretations*
- Table 4.16 *Item Misfit: Final Study with 52 Participants*
- Table 4.17 *Person Misfit: Final Study with 52 Participants*
- Table 4.18 *Pearson Correlations: SDS Sum scores and Scenarios*
- Table 4.19 *Total Variance Explained: Standardized Residuals*

## LIST OF FIGURES

Figure 2.1 *Patterns of Practice for Equity: A Working Definition*

Figure 3.1 *An Example of Scenario*

Figure 4.1 *Variable Map: Pilot Study*

Figure 4.2 *Cumulative Probability Map by Andrich Thresholds: Pilot Study*

Figure 4.3 *Variable Maps: Comparison between New Zealand and the U.S. participants*

Figure 4.4 *Variable Maps: Comparison between novice and experienced participant*

Figure 4.5 *Variable Map: Pilot Study*

Figure 4.6 *Variable Map: Final Study (57 Cases)*

Figure 4.7 *Cumulative Probability Map by Andrich Thresholds: Final Study (57 Cases)*

Figure 4.8 *Variable Map: Final Study (52 Cases)*

Figure 4.9 *Cumulative Probability Map by Andrich Thresholds: Final Study (52 Cases)*

Figure 4.10 *Construct Definition Map*

Figure 4.11 *Categorical Characteristic Curves (CCCs): Final Study (52 Cases)*

Figure 4.12 *Scree Plot: Standardized Residuals*

## CHAPTER ONE: INTRODUCTION

### The Problem

Educational inequality in achievement, resources, and opportunity to learn is a persistent issue across the globe, and is often associated with factors such as class, gender, race/ethnicity, linguistic backgrounds, ability, and immigration status (United Nations Educational, Scientific, and Cultural Organization [UNESCO], 2015). In the context of the U.S. specifically, students from non-dominant culture backgrounds (particularly Black and Hispanic students from low-income families, English language learners) tend to perform less well on subject-matter standardized assessments (The National Assessment of Educational Progress [NAEP], 2015), have lower high school graduation rates and higher dropout rates (The National Center for Education Statistics [NCES], 2013), and have lower rates of transitioning to college immediately following high school graduation (NCES, 2013).

In the past two decades, reform initiatives such as No Child Left Behind (2001) have aimed to close the academic achievement gap by adopting a test-based accountability system to hold students and teachers accountable for predetermined subject knowledge standards. This stringent accountability is promoted through competition among individuals (e.g., students, teachers), rewards and punishments in the race towards the standards. This dominant reform ideology regards education as the means to obtain individuals' economic wellbeing and a nation's economic productivity in the global economy. And, to address the educational inequalities in achievement is through "teaching everyone the same thing in the same way at the same time" (Sleeter, 2013, p. ix). This is different from viewing education as the end goal itself in a

democratic society where citizens are equipped to perceive, reason, and judge information and knowledge, have respect for diversity, and participate fully in deliberate dialogue and debate for the public good (Guttman, 1987/1999; Sleeter, 2013). The dominant educational reform ideology, which Sahlberg (2011, 2014) has called the Global Education Reform Movement (GERM), is not unique in the U.S. context but is widespread in a number of countries such as England, Australia, and New Zealand. However, whether reform through competition, standardization, and test-based accountability leads to the desirable change in classrooms or not remains highly debatable, not to mention the negative consequences such as narrowing curriculum, instruction and learning for traditionally marginalized and low-performing students (Cuban, 2013).

Many scholars have argued that there is a need to shift the focus from trying to “close” the achievement gap to looking into structural inequalities from historical, sociopolitical, economic, and moral perspectives (e.g., Ladson-Billings, 2006; Sleeter, 2013; Welner & Farley, 2010). Unequal educational opportunities, resources to learn, and learning outcomes are often tied closely to the intersecting systems of inequalities based on factors such as gender, class, race/ethnicity, parental wealth, language, and immigration status (Rice, 2015; Welner & Farley, 2010). Students of historically marginalized communities often encounter multiple barriers that are produced and reinforced by an unjust and inequitable system (Mitchell, 2013; Ladson-Billing, 1995a, 1999), and are likely to have less positive schooling experiences than their more advantaged peers.

Given the educational inequalities in opportunities and outcomes that are rooted in and perpetuated by the unfair and unjust system, many education scholars and practitioners argue that quality teaching must go beyond simply raising students' academic achievement. Instead, the purpose of education and teaching is to prepare students to be critical and responsible citizens with the capacity to participate fully in a democratic society (Banks, Cochran-Smith, Moll, Richert, Zeichner, & LePage, 2005). Approaches that fulfill this goal such as multicultural education (Sleeter & Grant, 1987), culturally responsive and relevant teaching (Gay, 2000; Ladson-Billing, 1995a; Villegas & Lucas, 2002), and teaching to promote social justice and challenge inequities (Cochran-Smith, 2010; Zeichner, 1993) have been developed to respond to the challenges in American classrooms.

Specifically, teachers must recognize their critical role in breaking the barriers of institutionalized inequalities by confronting the injustice and inequity in their teaching practice (Sleeter, 2013). Teaching with a social justice and equity goal is thus both practical and political – teachers not only encourage academic success and cultural competence in students but also empower them to critically examine social inequalities and to challenge the status quo. The assumption here is that teaching is inherently value-laden and never detached from the socio-political context.

Preparing and supporting teachers to enact teaching practice that responds to diversity, challenges educational inequities, and promotes social justice is a pressing yet daunting and complex task especially under the current education reform movement. Studies have documented the resistance, emotional struggle, and non-engagement observed in teacher candidates' learning about teaching with social justice and equity

goals and the multiple challenges encountered when enacting such teaching practice in school contexts (e.g., Achinstein & Ogawa, 2012; Buehler, Gere, Dallavis, & Haviland, 2009; Picower, 2009; Ukpokodu, 2011). While theories of equity-centered and socially-just teaching are substantial, we know relatively little about the factors that influence how teacher candidates understand and enact this kind of teaching practice in classrooms.

Therefore, we need to understand how to better prepare and support teacher candidates to teach all students (in particular, historically marginalized students) with a commitment to the goal of equity and social justice.

Instead of taking a technical, linear, and simplistic view of the impact of teacher education on teacher candidates' learning, some scholars argue that research on teacher education needs to take a complex view which enables new empirical questions to be posed (Cochran-Smith, Ell, Ludlow, Grudnoff, & Aitken, 2014a; Zeichner, 2005). For example, more research is needed to understand how teacher characteristics and prior knowledge interact with the arrangements of teacher education programs which then shape teacher candidates' experiences in the programs and their learning, and how these multiple interactions connect to their teaching practice (Zeichner, 2005). Given that teachers' classroom practice is an important outcome of teacher education, we need to know whether and to what extent teachers enact teaching practice for equity in the classroom. There is a need for the development of better measures to capture teachers' knowledge and the extent to which they enact equity-centered pedagogy in the classroom (Hollins & Torres Guzman, 2005; Zeichner, 2005). An instrument that captures teachers' equity-centered practice provides an essential piece of evidence for research that seeks to



improve program content and structure as well as building theory about how teacher candidates and novice teachers learn to teach for equity.

### **Purpose of the Proposed Study**

The purpose of this study is to develop a self-report type of instrument to capture the extent of complex teaching practice that is grounded in a theory of teaching that promotes equity and social justice. This study is part of a larger research program – Rethinking Initial Teacher Education (RITE) led by researchers at the University of Auckland in New Zealand and Boston College in the United States. The specific contribution of this proposed study to project RITE is developing an instrument to capture the extent to which teacher candidates report enacting complex teaching practice grounded in a theory of teaching that challenges inequities and promotes social justice. While project RITE’s current research sites are initial teacher education programs at the University of Auckland in New Zealand, this study is situated primarily in the U.S. Despite the many differences between the two countries due to the historical, sociopolitical, economic, and institutional contexts, the challenges in educational inequalities and opportunity to learn among historically marginalized students and the influence of the dominant education reform ideology on teacher education policy and practice are similar (Cochran-Smith, Piazza, & Power, 2013; Grudnoff & Ell, 2013). Therefore, the rationale for the development of such an instrument is shared.

Project RITE is informed by “complexity theory integrated with critical realism” (CT-CR) to understand the complex causal relationships between initial teacher education policies/practices, teacher candidates’ learning, and school students’ learning (Cochran-Smith, Ell, Grudnoff, Ludlow, Haigh, & Hill, 2014b). The overall goal of project RITE

is to develop an explanatory theory of teachers' learning during the critical period of initial teacher education that helps us understand the complex factors that influence whether, how and to what extent teacher candidates enact practice for equity and understand that part of their work as teachers is joining with others to challenge inequities (Cochran-Smith et al., 2014b). The theoretical platform provided by CT-CR offers a lens to look at teacher education and research on teacher education with a complex view that recognizes individuals, schools, and teacher education programs as complex systems. Moreover, these complex systems interact, change, learn, develop, and evolve over time in the larger context of intersecting systems of inequalities based on gender, class, race/ethnicity, socioeconomic status, language, and ability (Cochran-Smith et al., 2014a).

To identify teaching practice that contributes to diverse students' academic, social, emotional, civic, and critical learning, the project RITE team searched internationally to select research-informed syntheses or programs of research that reflected a complex view of teaching. Five syntheses and programs of research were selected, including three New Zealand Best Evidence Syntheses (Aitken & Sinnema, 2008; Alton-Lee, 2003; Anthony & Walsaw, 2007), the Teaching and Learning Research Project [TLRP] in the UK (James & Pollard, 2006), the Measurements of Effective Teacher in the U.S. (MET Project, 2013),<sup>1</sup> Te Kotahitanga Effective Teaching Profile (Bishop, Berryman, Cavanagh, & Teddy, 2009), and the Center for Research on

---

<sup>1</sup> The US-based Measures of Effective Teaching (MET) project includes several frameworks and instruments: the TRIPOD survey (Fergusson, 2001) to assess students' experience in the classroom, Framework for Teaching [FFT]–Domain 2: The Classroom Environment and Domain 3: Instruction (Danielson, 2007); Classroom Assessment Scoring System [CLASS] (Pianta, la Paro & Hamre, 2008); Protocol of Language Arts Teaching Observations [PLATO] (Grossman et al, 2011); Mathematics Quality of Instruction [MQI] (Hill et al., 2008) and UTeach Teacher Observation Protocol [UTOP] (Walkington et al., 2011). Given the accessibility of the frameworks/instruments, this study only examined Danielson's Framework for Teaching [FFT]–Domain 2: The Classroom Environment and Domain 3: Instruction (Danielson, 2007).

Education, Diversity, and Excellence's [CREDE] five standards for effective pedagogy (Dalton, 2007). Based on content analysis of these syntheses, the RITE team identified six interconnected principles of teaching practice that are associated with positive student learning outcomes broadly construed, and named the construct as teachers' enactment of practice for equity. I will elaborate the six interconnected principles of teaching, which are called "facets" later, in detail in the following chapter.

The construct of teachers' enactment of practice for equity is a complex one. This suggests that any instrument (or items) developed to measure a complex construct like this must reflect its complex nature. The overall purpose of the instrument that captures teachers' enactment of practice for equity based on their self-reports is intended to be formative. In other words, this instrument is used for diagnosing to what extent teacher candidates and novice teachers enact equity-centered teaching in the initial critical years of learning to teach and capturing the growth of teacher candidates along their journeys. Given the complex nature of the construct and intended purpose of the instrument, a novel approach – a combination of Rasch measurement theory (Rasch, 1960) and Guttman's facet theory (Guttman & Greenbaum, 1998) - was chosen (based on rationale that will be elaborated in the second chapter) to develop this instrument. The research question of this study is: Can the construct of teachers' enactment of practice for equity based on teachers' self reports be measured reliably, validly, and authentically by applying a novel approach to construct a Rasch-based scenario-style scale?

### **Significance of the Study**

As discussed earlier, educational inequalities and inequities are major problems in the U.S. and internationally. Many have called for the transformation of teacher

preparation to respond to the challenges, and in the past two decades or so, much work has been produced about the theories and practice of teaching and teacher education that challenge systemic inequity and promotes social justice. Preparing and supporting teacher candidates to teach culturally and linguistically diverse learners with commitment to social justice and equity is a pressing task. However, we know little about how the arrangements of preparation programs shape teacher candidates' learning, what and how factors other than the arrangement of programs shape teacher candidates' learning and experiences in and beyond the programs that are conceptualized and designed to challenge educational inequality and inequity. To better prepare and support teacher candidates' learning in and beyond preparation programs, we need to know what practice for equity is about, how we can capture the complex practice meaningfully and reliably, how the enactment of practice evolves and changes, and whether and how teaching practice for equity influences students' academic, social, emotional, critical and civic learning.

Given this research context, this study is significant for two main reasons. First, an instrument that reliably and meaningfully measures teachers' self-reports about their enactment of practice for equity would provide a new way to capture teachers' practice that provides evidence about the outcomes of teacher preparation. This would contribute to research that seeks to improve programs and/or to investigate how, why, to what extent, and under what conditions teacher candidates learn to enact this kind of practice, and the connection to student learning outcomes. Moreover, research and evaluation on teacher preparation programs with these purposes can provide more fine-grained information about how to improve teacher preparation programs. This can be a robust

alternative to the current dominant reform initiatives that often rely on stringent accountability mechanisms – causal linear logics and narrowly defined measures - to generate change (but not necessarily improvement).

Second, based on my preliminary research, the majority of instruments developed to measure attitude, belief, knowledge, or performance of equity-centered social justice teaching apply Classical Test Theory (CTT) as the approach to develop items and conduct analyses. While CTT is relatively easy to understand, commonly used, and provides useful summary information about participants' latent traits (e.g., teachers' attitudes, beliefs, or practice of equity-centered teaching), I could not find a scale that provides a meaningful authentic interpretation of what a score actually means in terms of teaching practice for equity and which allows a diagnostic analysis of teaching practice. To date, very few studies of instrument development construct scenario-like items by applying Rasch measurement theory and facet theory. In the following chapter, I argue that the combination of Rasch measurement theory and facet theory can offer a productive and systematic approach to develop an instrument to capture a complex construct, while addressing some of the disadvantages when using CTT. Thus, this study will contribute to the field of measurement in that it will explore this novel approach of scale development to capture a complex construct, teachers' practice for equity, in a meaningful and authentic manner.

Beyond the measurement community, this study can be of interest to the teacher education community – both scholars and practitioners – particularly those who are committed to transform teacher education to respond to educational inequity. In addition, this study can be of interest to researchers in the field of measurement as it provides a

rather novel case of instrument development using Rasch measurement principles and facet theory and the detailed processes involved.

### **Summary**

In this chapter, I introduced the challenges of educational inequalities among students from historically marginalized communities in many countries and the dominant education reform movement across the globe. Drawing from some scholars' existing work (e.g., Cochran-Smith, 2010; Ladson-Billings, 2006; Sleeter, 2013; Welner & Farley, 2010), I discussed the rationale for equity-centered socially-just teaching – its standpoint in addressing educational inequalities and inequities and fulfilling the democratic purposes of education. I have also offered brief discussion on challenges encountered in preparing and supporting teacher candidates to enact equity-centered socially just teaching practice in the classroom and the kinds of research needed to advance the field and better serve teachers and students. To understand whether and to what extent teachers enact practice for equity in the classroom as one outcome of teacher preparation programs, I suggested that there is a need to develop an instrument to capture teachers' equity-centered practice. Moreover, results of capturing teachers' equity-centered practice provide an essential piece of evidence for research that seeks to improve program content and structure as well as building theory about how teacher candidates and novice teachers learn to teach for equity.

Given this research need, I laid out the purpose of this study as to develop an instrument to measure teachers' self-reports about the extent to which they enact equity-centered practice in the classroom. I provided background information about the larger research program (Project RITE) that this study is part of. I also discussed how the

Complexity-Theory and Critical Realism lens adopted by project RITE informs the complex view of learning and teaching and the selection of five international syntheses and programs of research, which provide the conceptual base for the construct of teachers' practice for equity. I then specified the formative purpose of this instrument and the novel approach – combining Rasch measurement theory and facet theory – chosen to develop the scale given the measurement purposes. Lastly, I discussed the contributions of this study to the research on teacher education and the field of measurement.

In the following chapter, Chapter Two, a review of literature pertaining to this study is presented. The chapter first discusses the theory of equity-centered socially-just teaching and the six principles of teachers' enactment of practice for equity. Following the conceptual framework of equity-centered socially-just teaching is a review of existing instruments measuring different aspects (e.g., attitudes, efficacy, knowledge) of teaching with commitment to equity, social justice, and diversity. The review of existing instruments justifies the need for an instrument to measure teachers' practice for equity, and partly informs this study's choice of measurement theory. The third section of Chapter Two discusses and compares two prominent measurement paradigms, Classical Test Theory and Rasch measurement theory. The discussion also includes facet theory as a methodological approach to explore a content domain and its application in social science. I conclude the chapter by discussing the rationale for using the combination of Rasch measurement theory and facet theory in this study: why this approach is more suitable given the purpose of this instrument, how this approach can address some of the shortcomings observed in the existing instruments using CTT framework, and how this

novel approach guides the development of instrument to capture teachers' self-reports about their enactment of practice for equity.



## CHAPTER TWO: LITERATURE REVIEW

This chapter consists of four sections where I review some of the relevant literature that justifies and informs the design of this study. I first review theories of teaching and teacher education that promote equity and social justice. The conceptual model of teaching and teacher education for equity and social justice serves as the theoretical lens of this study as well as the theoretical foundation for the construct of teachers' enactment of practice for equity. Then, I review existing instruments that are designed to measure pre- and/or in-service teachers' attitudes, beliefs, awareness, self-efficacy, knowledge, skills, and/or competencies related to teaching for equity and social justice. The review of existing instruments justifies the need for an instrument to measure teachers' practice for equity. It also suggests that the commonly used approach of Classical Test Theory might not be suitable for the purpose of this instrument in addition to the approach's limitations. In the third section, I discuss the purposes, principles and assumptions, and mathematical models used by two prominent measurement theories, Classical Test Theory and Rasch measurement theory, and compare their advantages and disadvantages. The comparison of two measurement theories explains the choice of Rasch measurement theory for developing this instrument. I also introduce facet theory and the associated sentence mapping technique as a methodological approach to guide the item development procedure as well as the application of facet theory in social science. In the last section, I discuss the compatible view of Rasch measurement theory and face theory and previous successful applications of this novel approach to develop a scenario-style scale. I briefly discuss how this approach guides the process of instrument development in this study.

## **Teaching that Promotes Equity and Social Justice**

In response to the challenges posed by persistent inequities in the educational opportunities and outcomes between students historically marginalized by the system and their White counterparts, many scholars have proposed theories and practices of education intended to address issues of equity and social justice in K-12 settings (Dover, 2009). These include, for example, multicultural education (Sleeter & Grant, 1987), culturally responsive and relevant teaching (Gay, 2000; Ladson-Billings, 1995a, 1995b; Villegas & Lucas, 2002), socially just teaching (Cochran-Smith, 1999, 2010; Zeichner, 1993), democratic education (Howe, 1997), and critical pedagogy (Freire, 1970). While these theories of education have different roots and emphases, this body of work explicitly seeks to improve the opportunities and outcomes of historically marginalized learners (Cochran-Smith, Fiona, Grudnoff, Haigh, Hill, & Ludlow, 2016). Moreover, this body of education theories takes a critical socio-historical view of equity and equitable education (Cochran-Smith et al., 2016). This view of equity suggests that it is essential to acknowledge the unfair and unjust social and political system that contributes to and perpetuates educational inequalities in resources, opportunities, and learning outcomes. Therefore, teachers and teacher education must take the responsibility of challenging social structure and educational practice that perpetuates inequities, while acknowledging that teachers alone cannot eradicate educational inequalities and inequities. Lastly, critically-oriented education theories reject the idea that there is a set of teaching practices that are uniformly effective for all students regardless their cultural and linguistic identities and backgrounds, as well as the contexts of schools and classrooms.

The theoretical framework of this study, equity-centered social justice teaching (Cochran-smith, 1999, 2010; Cochran-smith et al., 2016), shares the same premises and purposes.

For the past decade or two, there has been an increasing interest and emphasis on social justice as the theme of teacher education; however, there is variation in how “social justice” is defined as a result of inadequate and ambiguous theoretical grounding (Cochran-Smith, 2010). Many have argued that without being explicit about the philosophical and political roots of social justice theory, teacher education for social justice and equity can easily become diluted and trivialized (Cochran-Smith, 2010; Grant & Agosto, 2008). In defining the construct of teachers’ enactment of practice for equity, this study adopts the framework of teacher education for social justice proposed by Cochran-Smith (2010). According to Cochran-Smith (2010), a theory of social justice for teaching and teacher education is “necessarily multi-perspectival, combining critical and democratic perspectives with commitments to anti-oppressive policies and practices” (pp. 449). A theory of social justice for teaching and teacher education recognizes and challenges systemic social inequities by attending to the dual dimensions of recognition of social groups and (re)distribution of goods (Cochran-Smith, 2010).

On proposing a theory of teacher education for social justice, Cochran-Smith (2010) argued that teaching for social justice and equity is not simply about what teachers know about subject content knowledge and what they do in the classrooms. Rather, teaching practice consistent with social justice and equity goals is defined by teachers’ knowledge, interpretive frame, methods, and advocacy (Cochran-Smith, 2010). First, teachers’ knowledge refers to not only the body of knowledge every teacher should know, but also the ability to “critique the very idea of knowledge and understand its

limitations” (Cochran-Smith, 2010, p. 455) – questioning the process of knowledge construction, who are considered legitimate knowers, and what can be known. This perspective of viewing teachers’ knowledge both recognizes and challenges the commonly accepted knowledge base, as well as includes knowledge that is historically marginalized. Second, an interpretive framework refers to the lens through which teachers make decisions, form relationships, view their own work, and support learning (Cochran-Smith, 2010). Some key interpretive frames for socially just teaching include seeing educators as agents of change, taking an asset-based approach (rather than a deficit lens) to multiple experiences brought by students, viewing one’s work through an inquiry stance, and understanding that teacher practice is inherently political. Third, methods refers to guiding principles for instructional approaches that are situated and contextualized, rather than best practices or techniques to be implemented faithfully for all students in any context. The guiding principles include: 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. Finally, the advocacy aspect of teaching practice refers to teachers’ explicit claims of their roles and responsibility of being advocates and activists with the commitment to recognize and challenge inequitable and unfair social and educational systems.

### **Principles of teachers’ enactment of practice for equity**

The theoretical framework of equity-centered and socially just teaching informs the construct of teachers’ practice for equity. The construct of practice for equity

comprises six interconnected principles of practice for equity, which were identified by the RITE research team members through qualitative content analysis of the five international syntheses and programs of research as mentioned earlier. Aligning with the framework of equity-centered social justice teaching, these six interconnected principles of practice for equity have been shown to improve broadly defined learning outcomes for marginalized, and all learners based on the empirical research reviewed in the five syntheses and programs of research (Cochran-smith et al., 2016; Grudnoff, Haigh, Hill, Cochran-Smith, ELL, & Ludlow, 2017). Moreover, the six interconnected principles of practice for equity are the general practice that reflects one's interpretive frame, view of knowledge, methods, and advocacy rather than universal teaching techniques about what a teacher should do in the classroom (Cochran-smith et al., 2016; Grudnoff et al., 2017). Viewing teaching and learning as complex and recognizing an important responsibility of teaching as to challenge inequities connects the theoretical framework and the six principles of teaching practice for equity. Table 2.1 presents the linkage between principles of practice drawn from the five international syntheses/research programs and the six interconnected principles (Grudnoff et al., 2017).

Table 2.1

*Linkages between the Five Syntheses and the Six Principles of Practice*

Theme	Characteristics/Principles/Domains on which the theme draws				
	BES*	TLRP	MET	CREDE	TK
1. Selecting worthwhile content and designing and implementing learning opportunities aligned to valued outcomes	1.1, 1.5, 1.6, 1.7, 2.2 3.2, 3.4, 3.7, 3.8, 3.9	1, 2	3b, 3a, 3c	4	
2. Connecting to students' lives and experiences	1.3, 2.1, 2.4 3.3, 3.5	3, 8	2a, 2b, 2c, 2d, 2e	2, 3	
3. Creating learning-focused, respectful and supportive learning environment	1.2, 2.3 3.3, 3.5	6, 7	3d, 3e,	1, 5	
4. Using evidence to scaffold learning and improve teaching	1.4, 1.8, 1.9, 1.10, 3.6	4, 5			
5. Taking an inquiry stance for professional engagement and learning	3.10	9, 10			
6. Recognizing and challenging inequities					Overall

\*BES – 1= General (Alton-Lee, 2003); 2= Social sciences (Aitken & Sinnema, 2008); 3= Mathematics (Anthony & Walshaw, 2007)

The six interconnected principles of practice that define the construct of practice for equity are presented as following (Grudnoff et al., 2017):

- 1) Selecting worthwhile content and designing and implementing learning opportunities aligned to valued outcomes: Characteristics include selecting content and setting learning goals and outcomes, designing and selecting learning opportunities, and implementing planned learning experiences aligned with valued outcomes;
- 2) Connecting to students as learners and their lives and experiences: Characteristics include identifying and recognizing students' home and community culture, and

- making connection between home and school cultures by planning and implementing linguistically and culturally sensitive curriculum, instruction, and assessment;
- 3) Creating learning-focused, respectful and supportive learning environments: Characteristics include fostering a caring, respectful, and inclusive learning environment, and orchestrating classroom procedures and physical space to facilitate collaborative learning;
  - 4) Using evidence to scaffold learning and improve teaching: Characteristics include designing assessment and using evidence to improve instruction and scaffold learning;
  - 5) Taking an inquiry stance for further professional engagement and learning: Characteristics include taking an inquiry stance for individual and collective learning, and collaborating with involved stakeholders to advocate for supportive and sustainable learning environments and professional culture;
  - 6) Recognizing and challenging classroom, school, and societal practices that reproduce inequity: Characteristics include recognizing and challenging one's own deficit thinking and theorizing historically marginalized students, and taking an agentic position in their practice and accepting professional commitment and responsibility.

The empirical evidence used in the five syntheses and programs of research suggests that these six principles of practice for equity are associated with positive student learning outcomes broadly defined to include academic, social, emotional, critical and civic learning. Moreover, these six principles are interconnected with each other and

it would be difficult to enact one principle without also enacting others. That is, it is the enactment of multiple principles as a whole rather than the enactment of one single principle that enhances student learning.

For example, “selecting worthwhile content and designing and implementing learning opportunities aligned to valued outcomes” (i.e., principle one) must rely on the understanding and recognition of students’ lives, cultural experiences, and prior knowledge (i.e., principle two). “Connecting to students as learners and their lives and experiences” (i.e., principle two) then facilitates the “creation of learning-focused, respectful, and supportive learning environments” (i.e., principle three). The enactment of these three principles also requires consistent use of assessment to scaffold learning and improve teaching (i.e., principle four), constant reflection and a strong commitment of professional responsibility and continuous learning (i.e., principle five). Lastly, the enactment of these five principles requires a strength-based framework to approach diversity brought by students, a critical lens to see the process of knowledge construction, and an agentic position to advocate for structural change (i.e., principle six).

The enactment of practice for equity is also highly contextualized. That is, the enactment of practice for equity is shaped by particular policy/political environments which are located in the larger interacting systems of inequality based on gender, race/ethnicity, culture, language, socio-economic status, immigration status, ability/disability. Therefore, the manifestation of practice for equity that is situated in specific contexts and time emerges in terms of patterns (teachers’ enactment of practice for equity). Figure 1 below illustrates the definition of patterns of practice for equity.<sup>2</sup>

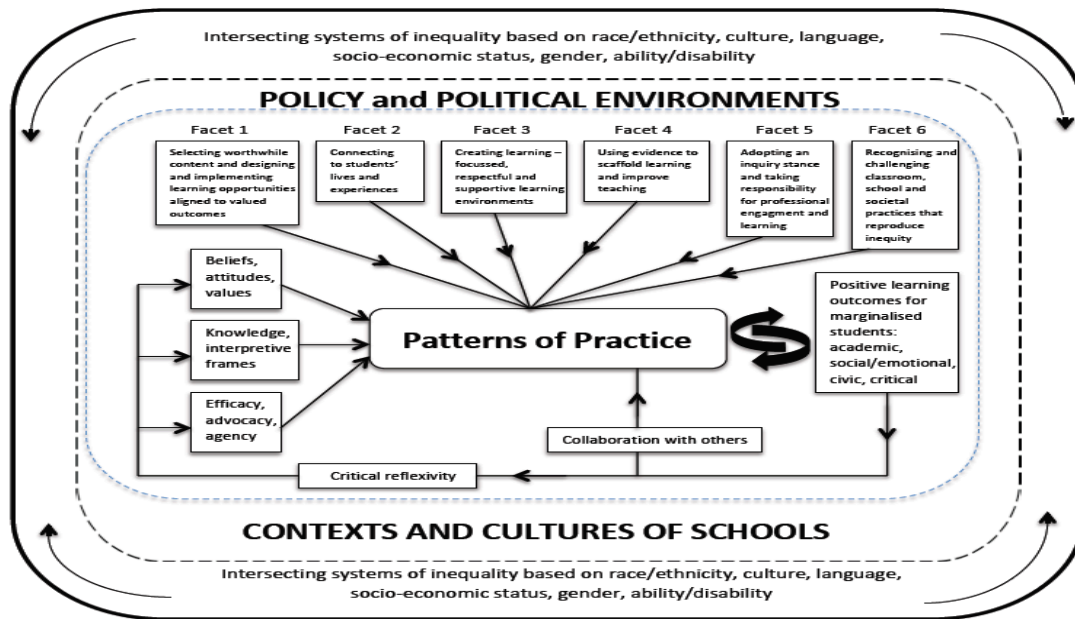
---

<sup>2</sup> This figure is reproduced from the paper presentation by the project RITE research team in the 2016 American Educational Research Association annual meeting in Washington D.C.



Figure 2.1

*Patterns of Practice for Equity: A Working Definition*



**Existing Instruments Measuring Aspects of Teaching for Equity and Social Justice**

The vast majority of existing instruments related to teaching for equity and social justice measure preservice and/or inservice teachers' attitudes (e.g., Baluch, Greig, Ponterotto, & Rivera, 1998), beliefs (e.g., Ludlow, Enterline, & Cochran-Smith, 2008), awareness (e.g., Henry, 1986), self-efficacy (e.g., Guyton & Wesche, 2005; Ritter, Boone, & Rubba, 2001; Siwatu, 2007), and knowledge (D'Andrea, Daniels, & Noonan, 2003) related to teaching diverse learners with equity and social justice commitment.

For example, the Cultural Diversity Awareness Inventory [CDAI] (Henry, 1986) assesses educators' attitudes, beliefs, and behaviors toward young children with culturally diverse backgrounds and young special needs children. The self-report CDAI consisting of 28 five-point Likert type items has been pilot tested with internal consistency reliability alpha of 0.9 and used in other studies (e.g., Rucker, 2004; Walker-Dalhouse & Dalhouse, 2006). Another example is the Learning to Teach for Social Justice-Beliefs

scale (Ludlow et al., 2008), developed by the evidence team of Boston College-Teachers for a New Era (BC-TNE) project. The scale consists of 12 five-point Likert scale items (i.e., 1 = *Strongly disagree*, 2 = *Disagree*, 3 = *Uncertain*, 4 = *Agree*, 5 = *Strongly agree*). Unlike most of the instruments discussed here, the Learning to Teach for Social Justice-Beliefs scale was developed using Rasch measurement principles. The instrument has reliability of above 0.7 and has been widely used on teacher candidates and graduates from Boston College, and teacher education programs nationally and internationally (e.g., New Zealand, Ireland) (Cochran-Smith, Ludlow, Ell, O’Leary, & Enterline, 2012).

Several instruments measure the self-efficacy and outcome-expectancy of teaching for culturally and linguistically diverse learners, including the Multicultural Efficacy Scale (Guyton & Wesche, 2005), the Self-Efficacy Beliefs about Equitable Science Teaching and Learning [SEBEST] (Ritter et al., 2001), the Culturally Responsive Teaching Self-Efficacy [CRTSE] Scale (Siwatu, 2007), and Culturally Responsive Teaching Outcome Expectancy [CROTE] Scale (Siwatu, 2007). Except for the Multicultural Efficacy Scale (Guyton & Wesche, 2005), the other instruments mentioned above were built on Bandura's (1977) self-efficacy construct which consists of two cognitive dimensions: personal self-efficacy and outcome expectancy.

Using the CRTSE and CROTE scales developed by Siwatu (2007) as examples, it is clear that culturally responsive teaching self-efficacy is defined as teachers’ beliefs in their confidence to execute specific culturally responsive teaching practices, while culturally responsive teaching outcome expectancy is defined as the degree to which teachers believe that their culturally responsive teaching contributes to positive classroom environment and students’ learning outcomes. The 40-item CRTSE and 26-item CROTE

both use a 100-point scale ranging from 0 (*not confident at all*) to 100 (*completely confident*). Both scales have Cronbach's alphas of above 0.95 and have been used in many research projects studying the impact of certain components of a teacher preparation program on the culturally responsive teaching self-efficacy of preservice teachers (e.g., Fitchett, Starker, & Salyers, 2012; Frye, Button, Kelly, & Button, 2010; Siwatu, 2009, 2011).

Overall, this body of research on preparing teachers for culturally and linguistically diverse students with a social justice and equity goal has focused on the learning and development of teacher candidates' cultural awareness, dispositions, attitudes, beliefs, and self-efficacy. While these factors are important predictors of teachers' practice in the classrooms, scholars (e.g., Sleeter, 2001) have been calling for research in teacher education to measure teachers' practice.

In addition to the common emphasis on factors predicting teachers' practice, a close examination of the theoretical foundations of existing instruments suggests that the political and critical aspects of teaching culturally and linguistically diverse learners with equity and social justice commitment are left out of most discussions (e.g., Baluch et al., 1998; D'Andrea et al., 2003; Siwatu, 2007). The only exception among the identified instruments is the Learning to Teach for Social Justice-Beliefs scale (Ludlow et al., 2008). Specifically, the authors see the primary consideration of teaching for social justice as "promoting pupils' academic, social, emotional, and civic learning and enhancing pupils' life chances, including challenging the structures, curriculum, labels, and school arrangements that limit or inhibit life changes" (Ludlow et al., 2008, p. 194). Moreover, the authors argue that teachers who enact social justice teaching function as

part of larger social movements and take an agentic position to advocate for their pupils (Ludlow et al., 2008). This element of “recognizing and challenging structural inequalities” is missing in the majority of studies that aim to get at aspects of teaching for social justice or equity. When the theoretical foundation does not reflect the political and critical aspects of teaching, the items of particular instruments are unlikely to capture the characteristics either.

For example, in discussing culturally responsive pedagogy in the development of the CRTSE and CRTOE scales, Siwatu (2007) regarded culturally responsive teaching as an approach to teaching and learning that incorporates students’ prior knowledge and previous experiences as a conduit to facilitate the teaching-learning process, to design classroom environments that are culturally compatible, to provide students with multiple approaches to demonstrate their learning, and to provide students with the knowledge and skills to function in mainstream society. However, there was no discussion about teaching that helps students to recognize, criticize, and challenge the current social inequities, to be critical of knowledge construction, and to share power of decision-making with students in the classrooms.

Another example of diluting the political aspect of teaching can be seen in the Multicultural Awareness, Knowledge, and Skills Survey – Teacher Form (MAKSS Form-T), a self-administered questionnaire consisting of 41 four-point Likert scale (1 = *Very Limited*, 2 = *Limited*, 3 = *Good*, and 4 = *Very Good*) items to measure teachers’ multicultural awareness (8 items), multicultural knowledge (13 items) and multicultural skills (20 items) (D’Andrea et al., 2003). While the authors referred to the principles of multicultural competency developed by the American Counseling Association (ACA),

there was no clear discussion of what multicultural competency for teachers is. When I examined the items developed to measure multicultural skills, I found that no items reflected the political aspects of multicultural education – for example, recognizing and critiquing the process of knowledge construction, reflecting and challenging the deficit lens of theorizing students from non-mainstream culture.

Some instrument development studies do not clearly define terms such as “equitable teaching” and “multicultural competencies.” For example, Ritter and colleagues (2001) discussed inequalities in the interactions between teachers and students who are girls, from low-income families, and ethnically and culturally diverse backgrounds. The authors also stated that the instrument was intended to “assess preservice teachers’ self-efficacy and outcome expectancy beliefs with regard to teaching and learning science in an equitable manner when working with diverse learners” (Ritter et al., 2001, p. 179). However, the authors never clearly defined the meaning of equity.

Some validated instruments that measure what teachers are able to do in the classroom do exist. However, these instruments are not grounded in the theories of equity-centered social justice teaching. I conducted a search of existing instruments published between 1971 and 2013 in the Mental Measurement Yearbook database by using keywords including “teaching” and “teachers.” The search results included the *Praxis* Performance Assessment for Teachers by the Education Testing Center (n.d.), the Scales for Effective Teaching (Kukic, Fister, Link, & Freston, 1989), Teacher Role Survey (Anderson & Maes, 1986), Teacher Evaluation Scale (McCarney & Cummins, 1986), Teacher Evaluation Rating Scale (Ysseldyke, Samuels, & Christenson, 1988), Teacher Performance Assessment (Soares & Soares, 1999), and Illinois Ratings of

Teacher Effectiveness (Blanchard, 1967). In addition, there are a number of validated instruments used to observe teachers' classroom practice such as Danielson's Framework for Teaching Evaluation Instrument (Danielson, 2014), the Classroom Assessment Scoring System [CLASS] (Center for Advanced Study of Teaching and Learning, n.d.), and Reformed Teaching Observation Protocol [RTOP] (Piburn & Sawada, 2000). Again, a close examination suggests that these instruments do not explicitly include recognizing and challenging equity in their frameworks.

All these assessments intend to measure teachers' subject content knowledge and/or what teachers are able to do in the classrooms (e.g., the *Praxis* Performance Assessment for Teachers by the Education Testing Center). While these instruments have sound psychometric properties and are useful tools for either summative (i.e., decision making) or formative (i.e., improvement) evaluation purposes, it is clear that there is a lack of attention to equity-centered teaching in these instruments – both in centering the commitment to equity in teaching practice and in conceptualizing what teaching for equity is like. Also, some identified assessments measuring teachers' pedagogical content knowledge and what they are able to do in the classrooms do not include teachers' interpretive frames and advocacy in teachers' practice, which is inconsistent with the theoretical framework of the present study.

In fact, similar critiques have been raised about research on preparing teachers for culturally responsive/relevant teaching. Critics argue that the political aspects of teaching culturally and linguistically diverse learners are often diluted in educational practice and research (Writer, 2008), and that culturally-responsive practice and research might not go far enough in terms of taking a critical stance to sustain cultural pluralism and to create

spaces for resistance to systematic oppressions (Paris, 2012). Some scholars also suggest that to transform teacher education, it requires “putting equity front and center” (Nieto, 2000) – that is, conceptualizing educational inequalities, understanding the role of teaching and teacher education in addressing educational inequalities, and defining the practice of equity (Cochran-Smith et al., 2015). The point I raise here is that if the critical and political aspects of teaching diverse students with a social justice and equity goal are not included at the stage of defining a construct, the instruments developed are very unlikely to reflect the critical and political role of teaching that promotes social justice and challenges inequities.

Last but not least, most of the identified instruments with the exception of the Learning to Teach for Social Justice-Beliefs [LTSJ-B] scale (Ludlow et al., 2008) used Classical Test Theory to theorize a construct, develop items and instruments, and generate scoring estimates for the instrument. Below, I briefly discuss three major limitations of the CTT approach that Rasch measurement theory (together with facet theory) can better address. The comparison of the two measurement theories is discussed in greater details in later sections of this chapter.

First, most of the studies reviewed here began by defining what the construct of interest was about (or not). Then the process of instrument development generally involved writing as many items as possible to sample the construct. This process usually required several rounds of review and empirical analysis to finalize the instrument (e.g., Guyton & Wesche, 2005; Marshall, 1996; Ritter et al., 2001). For example, in developing the Self-Efficacy Beliefs about Equitable Science Teaching and Learning (SEBEST), Ritter and colleagues (Ritter et al., 2001) initially drafted 195 items. It took

two review processes and a pilot study to eventually narrow down to the 34-item scale. This common approach to developing items and instruments is inefficient and unsystematic. Because, informed by CTT, items are replications of the construct and increasing the number of highly correlated and moderately difficult items will increase the total score reliability, decrease the standard error of measurement, and thus give more precise scoring estimates (Lord & Novick, 1968).

Second, CTT's scaling approach focuses on composite scores, and item estimates such as item difficulty and item discrimination are drawn from test-level statistics such as means and standard deviations (Hambleton & Jones, 1993). This means that, using CTT, test developers do not have the item-specific information to revise individual items when the test as a whole does not function well (e.g., low reliability, lack of discrimination due to too easy or too difficult test for the sample). Moreover, these item estimates are sample-dependent, and person scores are test-dependent – items can be more or less difficult depending on the sample respondents and vice versa (Hambleton & Jones, 1993). Along similar lines, the features of sample-dependent item estimates and test-dependent person scores reduce the utility of making comparisons between groups and measuring growth over time (Hambleton & Jones, 1993). With regard to this, the Rasch measurement models produce test- and sample-free estimates that allow objective comparison among individuals and groups and among items. The scaling approach of Rasch measurement theory also provides item-specific information if revisions need to be done to particular items to build a better scale.

Working from the CTT framework, the construct validity of an instrument is often an afterthought in the process of instrument development. While some efforts may be



expended upfront on defining a construct and determining whether items represent some aspects of the construct (e.g., through expert reviews) to ensure content validity, the steps taken to check the construct validity always come after finalizing the instrument. For example, to check the construct validity of the Teacher Multicultural Attitude Survey [TMAS] (Balugh et al., 1998), the authors tested the correlation between the TMAS and other existing scales (e.g., Multigroup Ethnic Identity Measure, Quick Discrimination Index racial) that measure the similar construct after finalizing the 20-item TMAS. Unlike CTT, Rasch measurement theory requires an a priori hypothesis about the estimated locations of items that define the construct of interest as a hierarchical continuum. The extent to which the calibrated items successfully capture the variation of the construct with reasonable spread as expected in the a priori hypothesis provides construct validity evidence (Wilson, Allen, & Li, 2006).

Overall, a review of 16 existing instruments measuring different aspects of equity-centered socially-just teaching revealed that the vast majority of existing instruments measure teachers' attitudes (e.g., Baluch et al., 1998), beliefs (e.g., Ludlow et al., 2008), awareness (e.g., Henry, 1986), self-efficacy (e.g., Siwatu, 2007), and knowledge (D'Andrea et al., 2003), rather than teaching practice. Moreover, among most of these identified studies, the political aspect of teaching - recognizing and challenging structural inequalities and taking an agentic position to advocate for students' learning – is often diluted or missing in the conceptual definition of the construct (e.g., Siwatu, 2007). In some cases, the idea of “equitable teaching” is not clearly defined (e.g., Ritter et al., 2001). Last but not the least, most existing instruments applied CTT to perceive a construct, develop items, and conduct statistical analysis. In addition to the shortcomings

of CTT briefly discussed earlier, the CTT approach is not suitable for the intended use of this instrument. That is, while these instruments using the CTT approach provide useful summary information about teachers' practice, attitude, or belief, a scale that allows formative assessment and a diagnostic analysis of status and change in teaching practice is needed. In the following section, I discuss the two measurement paradigms in greater details and argue that Rasch measurement theory is the more suitable and promising approach for this instrument.

### **Review of Measurement Theories**

In this section, I discuss two major paradigms of test theories – Classical Test Theory (CTT) (Lord & Novick, 1968) and Rasch measurement theory (Rasch, 1960/1980). For each framework, I discuss the fundamental definitions and assumptions made, the estimation procedure of key test statistics and parameters, the process of instrument development generally used for each framework, and the commonly reported psychometric results. I also compare the advantages and disadvantages of using CTT and Rasch. Then I present principles and ideas about facet theory and the associated sentence mapping technique (Guttman & Greenbaum, 1998). While facet theory is not a measurement theory per se, it is used by this study to guide the process of construct clarification and item development that is critical in applying Rasch measurement theory.

#### **Classical Test Theory (CTT)**

In the following section, I first discuss the fundamental concepts and assumptions of Classical Test Theory. I then explain the concepts of reliability, statistical methods used, and psychometric results reported when working from the CTT framework. I

conclude the section by summarizing the features and limitations of CTT regarding its approach to measuring a construct and the interpretation of psychometric results.

***The fundamental concepts of CTT.*** Classical Test Theory (CTT), also known as True Score Theory, is based on the concept that one's observed score ( $X$ ) is made up of two unobservable components: one's true score representing the latent construct of interest ( $T$ ) and error score or measurement error ( $E$ ) (Lord & Novick, 1968). The linear relations of observed, true, and error score is presented in the first equation:

$$X = T + E \quad (2.1)$$

To obtain a summary of a person's ability on an unobservable (latent) trait, a test is often used to sample the latent trait of interest. Ideally, if we can administer multiple tests to sample the trait and calculate the average score of all samples, we have a better picture of the individual's inner trait. While it is often not the case where multiple tests can be administered to measure the trait of interest, the idea of sampling a trait or a construct is useful in understanding the fundamental concept of CTT. Considering that an individual's latent trait of interest remains unchanged (i.e.,  $T$  is constant), the observed score ( $X$ ) through a single test is a chance variable with an unknown frequency distribution (Lord & Novick, 1968). The expected value of this frequency distribution, or the expected value of observed scores, is one's true score ( $T$ ). Based on the linear equation, the error of measurement ( $E$ ), also a random or a chance variable, is the difference between one's observed score and true score. The true score is a useful theoretical concept, and a convenient statistical device to allow the explanation of variance of observed scores and is never obtained (Lord & Novick, 1968). Given this linear and additive model then one's observed score is the sum of one's true score and

some systematic and random error. The true score is of real interest and the focus is to reduce the measurement error to obtain a more precise estimate of the true score. The following question is how can measurement error be reduced to have a more precise estimate of the true score under the CTT framework?

***Assumptions of CTT.*** In order to express the unknown parameters of measurement error and true scores based on the parameters of observed scores, some assumptions must be made to solve the equation (Lord & Novick, 1968). These assumptions are: a) the variance of observed score, true score, and error over all persons are finite and larger than zero (Equation 2.2); b) the responses to any two items (e.g., X1 and X2) are independently distributed; c) the correlation between the true score and the error random variables is zero (Equation 2.3); d) the measurement error on one test is uncorrelated with true scores on different tests (Equation 2.4); and, e) the measurement errors on different tests are uncorrelated with each other (Equation 2.5) (Lord & Novick, 1968; Lord, 1980). In addition, because measurement error is the difference between true scores and observed scores, the expected value of measurement error can be proved statistically to be zero given that the expected observed scores and true scores are equal to each other (Lord & Novick, 1968).

$$0 < \sigma_X^2 < \infty, 0 < \sigma_T^2 < \infty, \text{ and } 0 < \sigma_E^2 < \infty \quad (2.2)$$

$$\rho_{TE} = 0 \quad (2.3)$$

$$\rho_{T2E1} = 0 \quad (2.4)$$

$$\rho_{E1E2} = 0 \quad (2.5)$$

***Concepts of reliability and measurement errors.*** Classical Test Theory defines reliability as the proportion of the observed score variance that can be accounted for as

the true score variance. From the definitions and basic assumptions discussed above, the formula of reliability can be obtained through examining the variances and correlations among the observed, true, and error scores. Given the zero correlation between measurement error and true score (i.e., the third assumption shown in the third equation above), the variance of observed scores can be expressed as the sum of variance of true score and error scores in the population of participants (Equation 2.6) (Lord & Novick, 1968). Moreover, the squared correlation of X and T, which is the fundamental definition of test reliability, can also be expressed as the proportion of variance of observed scores explained as the true score variance (Equation 2.7) (Lord & Novick, 1968). It is important to note that test reliability under the CTT framework can never be one because the variance of measurement error is finite and larger than zero as specified in the first assumption.

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2 + 2\sigma_{TE} = \sigma_T^2 + \sigma_E^2 \quad (2.6)$$

$$\rho_{XT}^2 = \frac{\sigma_{XT}^2}{\sigma_X^2 \sigma_T^2} = \frac{\sigma_T^2}{\sigma_X^2} = 1 - \frac{\sigma_E^2}{\sigma_X^2} \quad (2.7)$$

In order to obtain the estimate of  $\rho_{XT}^2$  based on the known information (i.e., the observed scores and the variance of observed scores), CTT assumes strictly parallel test forms (Lord, 1980). Accordingly, the expected value of observed scores from one test form is equal to the expected value of observed scores from the strictly parallel test, which is equal to the true scores (Lord & Novick, 1968). The variance of observed scores from two test forms are equal to each other, so are the variances of random error variables. Based on this assumption, the squared correlation of the observed scores in the population and the true scores is equal to the correlation of the observed scores between the two strictly parallel test forms (Equation 2.8) (Lord & Novick, 1968). In other words,

the fundamental concept of reliability in CTT is derived from the strictly parallel tests. The unknown parameters, variance of the true scores and measurement errors, can then be expressed based on the variance of the observed scores and the correlation of the observed scores between the two strictly parallel test forms (Equation 2.9 and 2.10) (Lord & Novick, 1968).

$$\rho_{XT}^2 = \rho_{XX'} = \frac{\sigma_T^2}{\sigma_X^2} \quad (2.8)$$

$$\sigma_T^2 = \sigma_X^2 \rho_{XX'} \quad (2.9)$$

$$\sigma_E^2 = \sigma_X^2 (1 - \rho_{XX'}) \quad (2.10)$$

In addition, from the equation of the variance of measurement errors, we can then derive the equation for standard deviation of measurement errors, or standard error of measurement (Equation 2.11). It is important to note that the standard error of measurement is the same for each person's score (Lord & Novick, 1968). To address the earlier question of how to obtain a more precise estimate of true scores, the equation below (Equation 2.11) suggests that increasing test-retest reliability ( $\rho_{XX'}$ ) reduces the standard error of measurement, and in turn increases the accuracy of estimating the true scores (Equation 2.9). Again, this conceptualization of reliability is grounded in definitions of CTT (i.e., linear additive relations among observed, true, and error scores) and its assumptions.

$$\sigma_E = \sigma_X \sqrt{1 - \rho_{XX'}} \quad (2.11)$$

Furthermore, increasing the length of a test by adding items that are parallel measurements increases test reliability (Lord & Novick, 1968). Thus far, the linear equation of CTT (i.e.,  $X = T + E$ ) and the calculation of variance only represent one item as a test. In most cases of instruments measuring latent traits, a test or an instrument

often includes multiple items and is regarded as a *composite measurement* (Lord & Novick, 1968). The formula to calculate variances for the observed, true, and error scores are demonstrated by Lord and Novick (1968) and presented below.  $X$ ,  $T$ , and  $E$  represent the composite measurement,  $Y_i$ ,  $T_i$ ,  $E_i$  represent the individual components or items of the composite measurement, and  $n$  is the number of items in the composite measurement.

$$\sigma_X^2 = n \sigma_Y^2 [1 + (n+1)\rho_{YY_i}] \quad (2.12)$$

$$\sigma_T^2 = n^2 \sigma^2 (T_1) \quad (2.13)$$

$$\sigma_E^2 = n \sigma^2 (E_1) \quad (2.14)$$

Given the linear relationship of the three variance components ( $\sigma_X^2 = \sigma_T^2 + \sigma_E^2$ ) and reliability as the ratio of variance of the true and the observed scores, when variance of measurement error decreases, variance of true scores increases, and reliability increases. Also, when the number of items increases, the variance of true scores will increase much faster due to the factor of squared  $n$  than the variance of measurement error which only increase by the factor  $n$  (Lord & Novick, 1968). Again, it is important to note here that reliability increases when the length of a test increases, and the items added are supposed to “measure the same thing” (i.e., parallel measurements) (Lord & Novick, 1968, p. 95).

In measuring the internal consistency of an instrument, Cronbach’s  $\alpha$  and the Kuder-Richardson formula 20 (KR-20, for dichotomous items) as the lower bound on the reliability of a composite test are often used (Lord & Novick, 1968). The formula for Cronbach’s  $\alpha$  is presented below (Equation 2.15), where  $n$  is the number of items in a composite test,  $\sigma_X^2$  is the variance of the composite observed scores and  $\sum_{i=1}^n \sigma^2(Y_i)$  is the variance of all items. The formula of the variance of the composite observed scores is

presented below as well, which is the sum of the variance of all items and the covariance of each pair of items in the composite test. In the formula of the variance of the composite test,  $i$  and  $j$  indicate any pair of two individual items in the composite test. To increase  $\alpha$ , one would want to maximize  $\sigma_X^2$  specifically by maximizing the covariance element  $\sum_{i=1}^n \sum_{j=1}^n \rho(Y_i, Y_j) \sigma(Y_i) \sigma(Y_j)$  of the equation. This can be achieved by having variance of each item to be around 0.5, which gives the largest variance of the composite test. In other words, items that are not too difficult or too easy would get the most spread of responses from the participants and thus generate the most information (where information is defined as item or score variance). Additionally, we also want items to be positively correlated with other.

$$\alpha = \frac{n}{n-1} \left( 1 - \frac{\sum_{i=1}^n \sigma^2(Y_i)}{\sigma_X^2} \right) \quad (2.15)$$

$$\sigma_X^2 = \sum_{i=1}^n \sigma^2(Y_i) + \sum_{i=1}^n \sum_{j=1}^n \rho(Y_i, Y_j) \sigma(Y_i) \sigma(Y_j) \quad (i \neq j) \quad (2.16)$$

For binary items, KR-20 is used instead and the formula is presented below (Equation 2.17) where  $p_i$  is the probability of answering the item ‘correctly’ (e.g., answering 1 versus 0) and  $q_i$  is the probability of 1 minus  $p_i$  (Lord & Novick, 1968). In a way,  $p_i$  is also the item difficulty – number of participants out of total that respond to the item correctly (i.e., choose 1). Again, items with 0.5 difficulty give the highest reliability of the composite test too ( $\sum_{i=1}^n P_i q_i$  is the largest when difficulty is right at 0.5).

$$\frac{n}{n-1} \left( 1 - \frac{\sum_{i=1}^n P_i q_i}{\sigma_X^2} \right) \quad (2.17)$$

***Psychometric results using the CTT approach.*** The basic psychometric tools of CTT include item analyses (i.e., item difficulty and item discrimination), reliability



analyses (i.e., internal consistency), factor analysis, and validity analysis (e.g., content, construct, and discriminant). With regard to item analyses, item means would be used as indicators for rating-scale item difficulty (i.e., low means = harder item and high means = easier item). Item discrimination indicates how effectively an item functions to distinguish between examinees who are relatively high on the measure and those who are relatively low on the measure. Item discrimination for non-dichotomous items can be obtained through Pearson's  $r$  (i.e., Pearson product moment correlation coefficient, item total correlation). In calculating Pearson's  $r$  ( $r_{X_{ij}, Total}$ ) the specific item is included in calculating the total variance, and as a result Pearson's  $r$  is likely to be inflated. Therefore, the corrected item-total correlation would be used instead. It is "corrected" in the sense that to obtain the corrected item total correlation for an item, the specific item is taken out when calculating the total variance, which is usually preferred. The value of corrected item total correlations range between -1 and 1, and ideally we would like to see high positive corrected item total correlations. With regard to reliability analysis, Cronbach's  $\alpha$  is often used and the threshold of .7 is considered to be an acceptable level of reliability.

***Summary of the CTT framework.*** CTT's conceptual model and estimation of parameters are straightforward to understand. However, the assumptions of CTT are weak, meaning that it is easy for the data to meet the assumptions (Lord & Novick, 1968; Hambleton & Jones, 1993). Therefore, Classical Test Theory is widely used in construction of scales and surveys (Hambleton & Jones, 1993). However, there are certain features and limitations of CTT associated with the definitions and assumptions of the framework.

Working from the CTT framework, items in a test are considered to be “parallel measurements” of the same thing (i.e., replications of the construct). Ideally, items are to be highly correlated with each other (larger covariance between pairs of items) and should not be too easy or too hard to the sample participants (Lord & Novick, 1968). CTT uses test-based statistical models and focuses on the level of composite test scores. Item estimates such as item difficulty and item discrimination as well as reliability estimates are based on test-level statistics and are calculated using raw scores (e.g., mean, standard deviation, and variance) (Hambleton & Jones, 1993). A respondent’s performance or ability is also presented by a total raw score measured through a set of items in the instrument.

A feature related to the use of raw scores is that the estimates of items and person performance are sample-dependent and test-dependent under the CTT framework. If a test is administered to three groups of respondents with different levels (e.g., high, moderate, low) of performance or ability on the variable of interest, the resulting item statistics and reliability estimates could be quite different depending on who or which group responds to the test. In other words, the estimation of item statistics is not separated from persons’ responses, and thus items can be more or less difficult depending on who responds to them. Similarly, a person who takes three tests measuring the same construct with different difficulty levels (i.e., very easy, just right, and very difficult) might receive different person raw scores. This means that respondents’ scores that are supposed to represent their true characteristics on the variable depend on the difficulty of the test.

The test-dependent and sample-dependent feature in the models of CTT

framework implies that the difficulty of test items developed must match the ability level of sampled participants. A “mismatch” between an instrument and participants would not provide much useful information and the reliability estimate would be low.

Therefore, it is especially important to develop items that are just at the right difficulty level to the sampled participants. The test-dependent feature under the CTT framework also limits the utility of test scores such as comparisons among groups and capturing change over time (Hambleton & Jones, 1993). Specifically, the difference between two persons’ total raw scores (and assuming that the difference between two persons’ true scores on the variable stays constant) might be different depending on tests used. If the test is either too easy or too difficult for both participants, there may be no difference between two persons’ raw scores. Even if the test can detect the difference between the two persons, raw score difference can be small or large depending on the spread of item difficulties. This suggests that the interpretation of raw score differences when comparing individuals can be arbitrary.

Also, working from the CTT framework, there is no a priori hypothesis on how the calibrated items are expected to define a construct and how respondents’ of certain ability levels are expected to answer items with a certain difficulty level. The item estimates such as item difficulty (i.e., item means) are based on test-level statistics (i.e., composite scores) and are sample-dependent. This can be problematic to test developers when psychometric results suggest that the test does not function well for a particular group of respondents (e.g., low reliability, items lack of discrimination). Given the lack of a priori theory along with CTT’s emphasis on the composite score and its sample-dependent item estimates, the CTT approach and the psychometric results generated

through its statistical models do not provide useful information for revision on individual items.

Lastly, respondents' performances are presented in raw scores in CTT and are treated as linear measures with equal intervals. Measurement error is also assumed to be the same for all respondents' scores in the CTT framework. However, this is often not the case for respondents who are on the top or at the bottom, as measurement error tends to be high for these people given the little information available on them.

### **Rasch measurement theory**

In the following section, I first present Rasch theory's perspective of measurement. I then delve into the measurement principles required for developing a Rasch-based instrument, key features in the analytical models, a general instrument development process, and the use and interpretation of psychometric results. I conclude the section by comparing the CTT and Rasch measurement theory. Specifically, I emphasize how Rasch measurement theory addresses some limitations associated with the CTT approach, and I provide relevant Rasch-based instrument studies to illustrate the discussions.

*View of measurement – A Rasch paradigm.* Rasch measurement theory, developed by George Rasch (1960) and expanded by succeeding scholars (Andrich, 1978; Wright & Masters, 1982), is grounded in a rather strict view of and requirement for measurement. Measurement, according to Wright and Masters (1982), “begins with the idea of a variable or line along which objects can be positioned, and the intention to mark off this line in equal units so that distances between points on the line can be compared” (p. 1). For measurement to be useful in scientific research, measures about respondents'

latent trait must provide objective comparison, prediction, and/or diagnosis of respondents' current condition to capture improvement (Wright & Masters, 1982). Objective comparison and useful prediction can be achieved, only if the construction of variable provides a systematic and reproducible relation between persons' measures and calibrated items (Rasch 1966; Wright, 1967; Wright & Masters, 1982). This means that measures of persons' performance or ability do not depend on the difficulty of the test items, and item estimates are independent of who actually responds to the test (Wright, 1967). The quest for objective measurement is the cornerstone of the development of Rasch measurement theory (Andrich, 1989; Wright, 1980). This view of measurement offers a paradigm that is different from the CTT framework in conceptualizing a variable (a latent construct), constructing observations to capture the variable, and using the kind of mathematical function (or measurement model) to bring observations and idea of measurement together.

***Rasch measurement principles.*** There are several essential principles of Rasch measurement theory that guide the thinking of a variable, the kind of items to capture experiences related to the variable of interest, and modeling approaches (Ludlow, Matz-Costa, Johnson, Brown, Besen, & James, 2014; Rasch, 1960; Wright & Masters, 1982). First, a construct or a variable is a unidimensional attribute of respondents. That is, objects or respondents of measurement possess multiple characteristics or 'dimensions,' and measurement of a variable captures one characteristic of respondents that represents one dimension of them (Andrich, 1989; Ludlow et al., 2014; Wright & Masters, 1982). Second, this unidimensional construct can be conceptualized in the form of order (i.e., continuum) indicating "more" of "less" amount of the variable. Therefore, items

developed are intended to capture the variation (i.e., “more” or “less”) of the amount, strength, or capacity of some sort in persons. And, how well persons with a certain amount of the variable fare against items with a certain amount can be determined along an abstract hierarchical continuum for comparisons among persons and items (Andrich, 1989; Ludlow et al., 2014; Wright & Masters, 1982).

Third, items are spaced along the hierarchical continuum of the construct evenly or in a sensible way according to the hypothesized order (Andrich, 1989; Ludlow et al., 2014; Wright, 1980; Wright & Masters, 1982). Fourth, item discrimination is set to be equal across all items (Ludlow et al., 2014; Wright & Masters, 1982). Fifth, items are “stochastically independent” – “the probability of a certain answer to an item is unaffected by the answers given to the other item” (Andrich, 1989; Rasch, 1966, p. 93). Lastly, a Rasch measurement model is used for confirmatory purposes (Andrich, 1989; Ludlow et al., 2014). As Wright and Masters (1982) indicated, “The implementation of a variable requires the construction and observation of enough actual examples to confirm the expected order and document the basis for inference” (p. 5).

***Rasch measurement models and key features.*** The Rasch measurement model is a family of models including the dichotomous, partial credit, rating scale, binomial trials, and poisson counts models, just to name a few (Wright & Masters, 1982). The choice of model depends on how the data are collected and their intended use (Wright & Masters, 1982). The core of the Rasch models is the idea that the outcome of the encounter between an item and a person is governed by the product of the persons’ ability and the item’s difficulty (Wright, 1967). In other words, the Rasch models ask what a person’s

chances are of producing certain responses to an item given his/her ability and the item's difficulty.

This fundamental idea is presented in the probabilistic model below (see Equation 2.18). The probability of a person  $n$  choosing a response  $x$  (or a desirable/right answer) other than  $x - I$  (or a less desirable/wrong answer) for item  $i$  is a function of the person's parameter (i.e., ability or amount of the latent variable in the person)  $\beta_n$  and the item's parameter  $\delta_{ix}$  (i.e., difficulty of the item). If  $\beta_n$  is equal to  $\delta_{ix}$ , the probability of choosing a response  $x$  other than  $x - I$  for item  $i$  is 0.5. This probability increases (i.e., gets closer to 1) when the person ability becomes larger than the item difficulty and the difference between  $\beta_n$  and  $\delta_{ix}$  becomes larger than 0. This probability decreases (i.e., gets closer to 0) when the person ability becomes smaller than the item difficulty and the difference between  $\beta_n$  and  $\delta_i$  becomes smaller than 0.

$$\pi_{nix} = \exp(\beta_n - \delta_{ix}) / 1 + \exp(\beta_n - \delta_{ix}) \quad (2.18)$$

Here, the measurement unit of person ability  $\beta_n$  and item difficulty  $\delta_{ix}$  is a *logit* (i.e., taking log of the odds ratio, which is the ratio of the probability of getting an item right, or an event occurring, and the probability of getting the item wrong, or an event not occurring). A person logit is the natural log odds of a person succeeding on an item with zero logit difficulty; while an item logit is the natural log odds of a person with zero logit ability failing an item (Ludlow & Haley, 1995). A more "able" person has a positive logit (i.e., when the odds is larger than 1) and less "able" person has a negative logit (i.e., when the odds is less than 1). For items, a hard item has a positive logit and an easy item has a negative logit. In short, logits are interval level units of measurement as a result of conducting an exponential transformation of total raw scores (Ludlow & Haley, 1995).

And, using this measurement unit allows us to put persons and items on the same interval-level scale.

Moreover, based on Equation 2.18 above, the difference between a person logit and an item logit gives the probability of an event occurring (or getting an item correct for a dichotomous item). When a person has a one logit advantage over an item regardless of what the person's logit or the item's logit exactly are, the one logit difference gives the same probability of an event occurring or getting an item right. For example, a one unit of logit scale advantage can occur anywhere along the hierarchical continuum - between a person with a logit of one and an item with a logit of zero or between a person with a logit of two and an item with a logit of one. In other words, the estimates of difference between person ability and item difficulty ( $\beta_n - \delta_{ix}$ ) has an equal interval – an equal change in logit scale and equal probability of getting an item right (Wright & Masters, 1982). It is important to emphasize that the equal change does not mean that the change is constant – the larger the logit difference is, the probability of an event occurring increases, but the increase is not constant.

As mentioned earlier, from the perspective of Rasch measurement theory, measurement must enable objective comparison. This requires a scale representing a hierarchical continuum of a construct that is reproducible regardless of the items used and the respondents sampled. The mathematical formulation of the Rasch measurement models in which observations are brought together to generate parameters enables the idea of and requirement for objective measurement. This brings another two essential features of the Rasch measurement models: *inferential separability* and *sufficiency*.



In the joint probabilistic function of Rasch measurement models, due to the additive relationship between the person and item parameters, the estimation of item parameters and the estimation of person parameters can be separated from one another (Rasch, 1966). *Separability* or *inferential separation*, as Wright and Master (1982) suggested, means that “each parameter and its associated statistics appear as a separate multiplicative component in the modeled likelihood of a suitable set of data” (p. 8). When the estimation of the person and item parameters can be separated in the model, one can obtain the estimation of person parameters by just using the total raw score of persons without dealing with the item parameters (Rasch, 1966). Similarly, one can obtain the estimation of item parameters by using the total item raw scores without dealing with the person parameters (Rasch, 1966). In other words, total raw score of the items are *sufficient estimators* to obtain item parameters and total raw score of the persons are *sufficient estimators* to obtain person parameters without knowing the specific patterns of responses (Rasch, 1966). As Wright (1980) pointed out, “When a sufficient estimate exists, it extracts every bit of knowledge about a specified feature of the situation made available by the data as formalized by the chosen model” (p. xii).

The features of *separability* and *sufficiency* are necessary conditions for *specific objectivity* (Rasch, 1966; Wright, 1967; Wright & Masters, 1982). The idea of *specific objectivity* (Rasch, 1966) suggests that the location of items (or how items are calibrated) along the hierarchical continuum of variable with measurement unit in *logits* are independent of persons responding to the items, and that the locations of persons along the scale are independent of the items to which they respond (Andrich, 1989; Wright, 1967, 1980). When *specific objectivity* is established in the measurement model, the

estimates of items or persons are non-arbitrary – a person’s estimate is not due to the particular test he/she takes and a test is not more or less difficult because of the respondents who take it. This is also what Wright (1967) called “sample-free test calibration and test-free person measurement” or measurement invariance (Wright & Master, 1982).

***Implications for instrument development and use of psychometric results.*** Given the goal of invariant comparisons and a rather strict set of model requirements, with the Rasch models, data should be carefully collected to fit the model (Smith & Andrich, 2005). This is achieved by having a clear understanding of the theoretical construct, conceptualizing the construct as a hierarchical continuum (i.e., what it means to possess low or less amounts to high or greater amounts of the construct), and carefully crafting items that capture the levels along the hierarchical continuum of the construct (Wilson, 2004). This process of careful construct clarification and item development is intended to control item biases (which may cause guessing or unequal discrimination) and ensure that characteristics of the construct contribute to item difficulty which influences persons’ responses. This process of developing the theoretical rationale and rank-ordered items according to the theory is the hypothesis that one must know and make prior to the test administration and data analysis (Andrich, 1989; Ludlow et al., 2014; Smith & Andrich, 2005; Wilson, 2004; Wright, 1980). And, the Rasch models serve as a confirmatory test of whether the data fits the model rather than a post-hoc analysis (Andrich, 1989; Ludlow et al., 2014).

The key psychometric results reported when using a Rasch model include the Rasch-Andrich “variable map” to check whether the data confirms the model, person and

item separation and reliability measures, category characteristic curves (CCCs) to check the distribution of response categories, and item and person fit indices (Smith, Linacre, and Smith, 2003). Because the Rasch measurement framework requires a priori theory of the variable, the psychometric information based on the empirical data provides a means of modifying the instrument. The information also provides the construct validity evidence to whether the calibrated items confirm the hypothesized order and define the construct.

*Variable map.* The variable map is a graphic representation of items' and persons' estimates along the hierarchical continuum of the construct. This hierarchical continuum is an interval-level scale with measurement units in *logits*. The variable map serves three main purposes. First, given that the variable map represents the construct of interest as a hierarchical continuum, it shows whether the items based on test data confirm the hypothesized order of the items (Wright & Masters, 1982). When empirical data do not confirm the hypothesized order of items, further clarification of the construct might be needed or particular items might be dropped or revised. Second, one would also want to see whether the calibrated items are sufficiently and reasonably spread out to define the increasing levels/directions of the variable (Wright & Masters, 1982).

In addition, the variable map provides information about whether items and people are on target with each other and how items or participants spread along the continuum. When persons and items are on target with each other, items are not too difficult or too easy for the sampled participants. When there is less gap between the locations of items and persons along the variable scale, the estimation of item and person parameters tend to be more precise with a smaller standard error of measurement.

Therefore, discrepancies between item and person locations along the continuum provide information for further item revision (i.e., make items more difficult or easier, or spread more evenly along the scale). Lastly, a person's location on the variable map indicates the person's level of construct at the time of measurement. This can be used as a baseline to identify progress made and to interpret the meaning of moving from one location to another along the construct continuum.

*Separation and reliability measures.* Similar the concept of Cronbach's alpha and KR-20 discussed earlier in the CTT section, Rasch separation reliability is also an estimate of the ratio of true measure variance (i.e., observed measure variance – error variance) to observed measure variance (Linacre, 1997; Wright & Masters, 1982).

Cronbach's alpha, which relies on the variance analysis of raw scores, generally produces higher reliability estimate; while Rasch separation reliability, which uses standard errors in estimating reliability (See Equation 2.19), usually underestimates it (Linacre, 2016).

Rasch separation and reliability of person and item measures have difference implications. Person separation and reliability indicates the degree to which the instrument is sensitive enough to reliably differentiate high-ability from low-ability individuals (Linacre, 2016). Ideally, the person separation index should be greater than 2 and person reliability should be greater than 0.8 (Linacre, 2016). Person separation and reliability can be influenced by factors, including the number of items, the number of response categories per item, the extent to which items and sampled respondents are on target with each other, and sample ability variance. Item separation and reliability indicates how well items are reliably separated to define the hierarchical continuum of the construct. Ideally, item separation index should be greater than 3, and item reliability

should be greater than 0.9 (Linacre, 2016). Item separation and reliability can be influenced by sample size, a range of item difficulty cover, and model fit (Linacre, 2016).

$$\text{Rasch Reliability} = 1 - \frac{\sum(\text{Standard Error})^2/N}{(\text{Standard Deviation of the Person Measures})^2} \quad (2.19)$$

*Category Characteristic Curves (CCCs).* Category characteristics curves present the probability of responding to each response category for any respondent on any item (Wright & Masters, 1982). The CCCs are often used to check whether each response option functions as intended – whether each response option is used (i.e., a minimum of 10 observations per category) and has about equal probability of response (Wolfe & Smith, 2007b). The category threshold measures, which reflect the scale value (e.g., 1 to 5 if a 5-point Likert scale is used), should follow an order (i.e., low to high value). The threshold measures that do not follow the ordered pattern indicate that some response categories are used more often than others. In this case, collapsing categories or revising the language of the less used categories might be needed.

*Goodness-of-fit.* Rasch fit analyses provide a standardized person-by-item residual between each person's observed response to an item and the expected response based on the model (Wright & Masters, 1982; Ludlow, 1986). Both unweighted and weighted model fit statistics for items and persons are used to identify outlier and inconsistent responses. The unweighted fit statistic, which is the mean square of standardized residuals between the observed and expected responses, is used to identify outlier responses across the entire set of participants for each item and the entire set of items for each person (Wright & Masters, 1982). While various cutoffs are used, the reasonable range for the unweighted mean square is between 0.6 and 1.4 and any value larger than 1.4 (for rating scale) indicates that at least one highly unexpected response has

occurred (Wolfe & Smith, 2007b). The weighted fit statistic, which is the mean square of standardized residuals weighted by the variance of test information, looks at the consistency of responses over the entire set of items for each person (a person-level fit statistic) and across the entire set of persons for each item (an item-level fit statistic). A threshold of 1.4 is often used for the weighted fit statistics (Wolfe & Smith, 2007b).

For the development of a new instrument, a more liberal threshold of 1.2 or 1.3 for both the unweighted and weighted fit statistics is recommended to avoid missing potential problems in the initial stage (Ludlow et al., 2008; Ludlow et al., 2014). Both weighted and unweighted mean square residuals can also be transformed into approximate t statistics. The absolute value of a t statistic that is larger than 2 (i.e.,  $t > |\pm 2|$ ) or 3 (i.e.,  $t > |\pm 3|$ ) depending on sample sizes indicates outlier or inconsistent responses (Wright & Masters, 1982). With a larger sample size, t statistics are often large due to the decrease of standard error and thus requires a higher threshold (i.e., 3 rather than 2). A positive and large magnitude of residual and/or t statistics indicates unexpected high or low responses (Wright & Masters, 1982). Further investigation is needed to understand what factors may contribute to the misfit and the kind of revisions needed.

***Comparison of Classical Test Theory and Rasch measurement theory.*** As mentioned above, differing test theories represent different paradigms for considering the important issues/priorities in measurement, conceptualizing a construct, developing the kind of observations to capture the variable, and building the kind of measurement model that fulfill the priorities and addresses technical issues (e.g., error) in measurement. In the sections that follow, I compare Rasch measurement theory to CTT. I suggest that

while the assumptions of CTT are easy to meet with test data and the estimation of parameters are straightforward to understand and interpret, there are several shortcomings associated with this approach that can be addressed by applying Rasch measurement theory.

First, the two measurement paradigms have different views on what a variable or a construct of interest is and what it means to develop items that provide person measures. With the CTT framework, items are considered to be replications of the construct – this means that items should be highly correlated with a moderate level of difficulty for the sampled participants (Lord & Novick, 1968). Item estimates (e.g., item difficulty, discrimination) are based on raw scores of some sort (e.g., item means and standard deviation) and person measures are individuals' total raw scores. Item estimates and correlation coefficient are used to determine whether an instrument functions well or not. However, as I will discuss shortly, CTT's reliance on the composite raw scores has some limitations. In contrast, Rasch measurement theory regards a construct as a hierarchical continuum and requires the development of items to capture the variation within a construct. That is, a priori theory through a careful construct mapping process must be determined prior to the collection and analysis of the empirical data (Ludlow et al, 2014; Wilson, 2004; Wright & Masters, 1982). Instead of using raw scores for both item and person estimates, a log transformation is done to produce calibrations of items and measures of persons (Wright & Masters, 1982). The order of calibrated items based on empirical data is then checked against the a priori theory. When data confirms the model, the variable map gives the evidence on construct validity.

Second, as described above, with the CTT framework, item estimates are calculated using raw scores and comparison among participants relies on person total raw scores on the test. Also, the standard error of measurement is the same for all respondents' scores. Given that test data are usually collected through a Likert-type ordinal scale, raw scores do not have the property of linearity (Wright & Masters, 1982). This means that for example a ten-point score difference in the lower or higher end of the score distribution might be interpreted differently from the ten-point score difference in the middle of the test score distribution. The distance between participants at either end of the variable is shorter than the distance if the item difficulties shift to target these participants (Wright & Masters, 1982).

Unlike CTT, Rasch models are probabilistic models, which allow the transformation of ordinal raw data (i.e., total personal score or item score) into an interval measurement scale with the measurement unit in *logits* (Smith & Andrich, 2005; Wright & Masters, 1982). The estimation of both person and item locations are placed on this measurement scale with equal intervals (Wright & Masters, 1982). The comparison between person ability and item difficulty also follows an interval-level scale. This means that a one logit difference between persons and items gives the same probability of answering the item correctly regardless of the locations along the variable scale.

The non-linearity of raw scores in CTT framework becomes particularly a problem when the same test is administered to participants of different levels or the same participants take a sub-set of test with different difficulty levels (Wright & Masters, 1982). Item estimates, say item difficulty, can be very different depending on the level of sample participants. Likewise, person measures (i.e., person total raw scores) can vary



depending on the difficulty of tests the participants take. This is then perhaps the most critiqued aspect of CTT: item estimates are sample-dependent and person measures are test-dependent. A scale and its psychometric properties that are dependent on samples is not particularly useful for comparing performance across groups or measuring change over time (Hambleton & Jones, 1993). When a scale is not ‘fixed’, the result of comparison becomes relative.

Rasch measurement models address this in the estimation process where person and item raw scores are sufficient statistics for estimation of person and item parameters respectively and either parameter can be estimated independent of the other (Rasch, 1966; Wright & Masters, 1982). These features then allow the objective comparison of items to be independent of particular individuals who respond to the items, and comparison of persons to be independent of items that are used. Assuming that the data fit the model (i.e., test items have good distributional characteristics and fit), the invariance of items is useful to diagnose abilities of persons along the scale, provide meaningful interpretation of person location in relation to the construct, and measure differences across groups and change over time (Hambleton & Jones, 1993; Wright & Masters, 1982).

Lastly unlike the CTT framework’s focus on the composite scores, Rasch measurement theory, which is an “individualistic approach to item analysis” (Rasch, 1966, p.89), provides analyses of items (e.g., item fit statistics, responses within an item). These analyses are based on empirical data that can be used to check whether the locations of calibrated items confirm the hypothesized order and to provide item-specific information for further revision. This gives Rasch measurement theory the power of

revising targeted items to build the best possible scale to capture the variation of a construct. This is not possible when using CTT, as item estimates are based on test-level statistics which are sample dependent, and thus revising items for a better functioning instrument can be like hitting a moving target.

*Application of Rasch measurement theory.* Rasch measurement theory has been used to measure behavioral constructs such as rehabilitation outcomes (Coster, Haley, Ludlow, Andres, & Ni, 2004; Coster, Ludlow, & Mancini, 1999), knowledge (Gable, Ludlow, McCoach, & Kite, 2011), and psychological constructs/traits such as conformity to masculine norms (Ludlow & Mahalik, 2001), confidence and anxiety for teaching economic literacy (Ludlow, Rollison, Cronin, & Wallingford, 2012), and engagement in later life activities (Ludlow, Matz-Costa, Johnson, Brown, Besen, & James, 2014).

Rasch measurement theory has increasingly been used in educational survey research (e.g., Evans, Brauchle, Haq, Stecker, Wong, & Shapiro, 2007; Fulmer, 2014; Funk, Fox, Chan, & Curtiss, 2008; Waugh, 2003) and a few applications are specifically related to teaching such as teachers' beliefs in learning to teach for social justice (Ludlow et al., 2008), teachers' assessment for learning classroom practices (O'Leary, Lysaght, & Ludlow, 2013), science teaching self-efficacy (Shireen Desouza, Boone, & Yilmaz, 2004), and teachers' attitudes toward homeless students (Brown, 2012).

For example, Ludlow and colleagues (2008) applied Rasch measurement theory as the framework to construct the Learning to Teach for Social Justice - Beliefs (LTSJ-B) scale. The unidimensional construct – learning to teach for social justice beliefs – is conceptualized as a hierarchical continuum ranging from lesser or weaker to stronger commitment to teaching for social justice. The final scale includes 12 five-point Likert

scale items (1 = *Strongly Disagree*, 2 = *Disagree*, 3 = *Uncertain*, 4 = *Agree*, and 5 = *Strongly Agree*). Each item is a statement representing central ideas of teaching for social justice, and the 12 items capture the variation of the construct. That is, some items are easier to endorse by most of teachers and some are harder to endorse. In their study, the Rasch rating scale model is used assuming that the difficulty of choosing a higher response category (e.g., *Disagree* versus *Strongly disagree*) is in order and the same threshold estimates apply to all items. Multiple administrations of the LTSJ-B scale have produced consistent psychometric results; construct validity of the scale (i.e., items capture the continuum of the construct as hypothesized) was established through empirical data (Ludlow et al., 2008). The LTSJ-B scale has also been used to capture differences in the beliefs of teacher candidates across cohorts and to measure change in beliefs within the same cohorts (Enterline, Cochran-Smith, Ludlow, & Mitescu, 2008). Moreover, as the authors emphasized, the scale is invariant (i.e., the hierarchical order of the items stays the same) across administrations with different cohorts, the same cohorts at different time points, and participants in other cultural contexts (Enterline et al., 2008).

Another example is an instrument developed to measure engagement in four different later life activities by Ludlow and colleagues (Ludlow et al., 2014). Guided by Rasch measurement theory, engagement as a unidimensional construct is conceptualized as a continuum ranging from none or low to higher levels of engagement and it is reasonable to consider that individuals can be located along this continuum. Applying a novel approach combining Rasch measurement theory and facet theory (detailed discussion on facet theory, its application alone and with the Rasch principles are presented below), nine five-point Likert scale scenario-style items for each of four

activities (i.e., work, caregiving, informal helping, and volunteering) were constructed to capture the variation of the engagement construct. Participants were asked to compare themselves with the situation described in each scenario item and rate whether they were *'much more engaged'*, *'more engaged'*, *'about as engaged'*, *'less engaged'*, or *'much less engaged'* than the situation in each item. It was expected that items capturing higher level engagement would be harder for most of participants to select *'about as engaged'* or above; while items capturing lower level engagement would be easier for most participants to select *'about as engaged'* or above. The Rasch rating scale model was used to analyze the data collected through the Likert scale items. Empirical results through the pilot study and full-scale administration suggest sound reliability and construct validity of the four scales (Ludlow et al., 2014). The construct of engagement measured through the scenario items was shown to be invariant across the four different activities/roles (i.e., the locations of calibrated items are the same for all four scales). Most importantly, to construct the best possible scale with scenario items that not only capture the variation of the construct but also have equal spread along the continuum of estimates, a follow-up study was conducted and successfully illustrated the scenario revision process and the final scale (Ludlow, Matz-Costa, & Klein, In press).

The examples above suggest that when empirical data confirm the model, instruments constructed by using Rasch measurement theory have the property of measurement invariance that allows objective comparison across groups and measuring change over time. The measurement framework also has the capability of allowing revisions to particular items in order to build the best scale possible. While measuring changes or revising the scale is not within the scope of this dissertation, it is the ultimate

goal of this instrument to be able to capture differences across and growth within cohorts of participating teachers. And to do so, a scale that provides useful and meaningful interpretation of the construct is necessary. Therefore, following the insights from previous examples, Rasch measurement theory is most suitable in developing this instrument.

### **Facet theory (FT) and the sentence mapping technique**

***Basic elements and definitions.*** According to Borg and Shye (1995), facet theory (FT) is a methodological approach that involves a design procedure to systematically explore a content domain of interest, data analysis, and correspondence hypotheses between the priori theory of the content domain and the empirical results. As a methodology in social science, facet theory is used to investigate and build a theory about a domain of interest through a systematic inquiry (Borg & Shye, 1997).

A facet is defined as “a set of attributes (variables) that together represent underlying conceptual and semantic components within a content universe” (Guttman & Greenbaum, 1998, p. 17). In other words, facets are a set of characteristics that are mutually exclusive and conceptually distinct from each other and are used to structure or represent an unstructured universe of content (Borg & Shye, 1995; Guttman & Greenbaum, 1998). Through *facitization* of a construct, it enhances the level of conceptual clarity and control in investigating a construct that contains an infinite number of items in the universe of content (Borg & Shye, 1995). It is important to note that the role of facets is not to “define a universe, but rather to structure a given unstructured universe” (Borg & Shye, 1995, p. 32). Within facets, there are elements, also called *structs*, which define the different values or levels that describe the variations within a facet (Guttman &

Greenbaum, 1998). The combinations of multiple facets' *structs* are then called *structuples* (Borg & Shye, 1995; Guttman & Greenbaum, 1998).

To facilitate the facet design, the sentence mapping technique is one of the basic features of FT in that it “facilitates theory construction in an explicit and systematic way of coordinating formal concepts and a language with informal ones” (Guttman & Greenbaum, 1987, p. 15). A mapping sentence consists of two main parts: a formal part drawn from facets and an informal part comprising the phrases linking facets together and providing the facets a context (Borg & Shye, 1995; Guttman & Greenbaum, 1987). In other words, a mapping sentence is a “lexical tool to illustrate the structural relationships among the facets” (Ludlow et al., 2014, p. 131) in a meaningful substantive context. Perhaps another way of seeing the sentence mapping technique is that it is a systematic approach to constructing an item rather than a “blindfolded approach” (as what is commonly seen in most instrument development processes). As Guttman and Greenbaum (1987) suggest, “constructing a mapping sentence forces the researcher to identify and explicate simultaneously the theoretical constructs of the research together with the kind of observations needed to test it” (p. 16). It can help avoid the lack of a clear connection between a construct definition and an item, minimize the involvement of other construct dimensions in the items, and ensure a greater level of content validity of items (Guttman & Greenbaum, 1987). Some suggest that the sentence mapping technique can be a very powerful approach to develop a scale measuring a social justice related construct (Vuilleumier, Klein, & Ludlow, 2015).

***Application of facet theory.*** As discussed earlier, facet theory as a methodological approach comprises a systematic procedure for exploring a complex social phenomena,

data analysis, and correspondence between the hypothesis and the empirical data. Facet theory has been applied to explore the underlying structure of complex domain such as motivations (Bilsky & Schwartz, 2008), workplace adaptive skill (Cronshaw & Jethmalani, 2005), teachers' professional ethics in Israeli context (Fisher, 2013), and organizational expectations of the novice teacher (Friedman, 2004, 2006). Among these studies, facets and elements within each facet are often identified based on existing social theories. The sentence mapping technique is then used to construct items to explore the underlying structure of the domain (e.g., Cronshaw & Jethmalani, 2005; Fisher, 2013). Research hypotheses on how facets and different combinations of elements within each facet (i.e., *structuples*) might influence participants' responses to each item are posed. Data analysis techniques such as smallest space analysis (SSA) or multidimensional scaling are often used to see whether the empirical data confirm the hypotheses (e.g., Fisher, 2014; Maslovaty, Marshall, & Alkin, 2001; Rettig & Leichtentritt, 1999). While many of these studies used facet theory and the sentence mapping structure to develop items to explore a domain of interest, the purpose of these items was to investigate the underlying structure of a domain (e.g., teachers' professional ethics) that consists of multiple variables or constructs.

### **Putting it together: Combining Rasch measurement theory and facet theory**

While most studies applying facet theory take on all three elements of facet theory (i.e., facet design, SSA as the data analysis technique, and correspondence hypotheses) (Bilsky & Schwartz, 2008; Cronshaw & Jethmalani, 2005; Fisher, 2013, 2014; Friedman, 2004, 2006; Maslovaty et al., 2001; Rettig & Leichtentritt, 1999), two recent studies adopted only the design aspect of facet theory and used the Rasch measurement models

as the analytical approach (Ludlow et al., 2014; Randall & Engelhard, 2010). The Rasch principles and the premises of facet theory reflect a comparable view towards investigating a construct and building an instrument that captures the construct in a meaningful and authentic manner.

First, both emphasize the conformity between the theory/design and the empirical evidence observed in the data. Moreover, Rasch measurement theory and FT both perceive a construct as a hierarchical continuum. Specifically, based on Rasch measurement principles, items must spread uniformly along a hierarchical continuum from easy-to-endorse to hard-to-endorse. Similarly, FT also considers each facet as having elements or variation within it (i.e., *structs*) either in terms of quantitative range (i.e., low to high) or qualitative content.

FT is particularly useful in developing Rasch-based instruments in that the design aspect of facet theory enhances the level of formality and better fulfills the principles of a Rasch measurement model. The facet approach breaks down a construct of interest into related variables or characteristics (i.e., facets) and items are developed explicitly to address these characteristics as well as variation within it. The combinations of *structs* from multiple facets (also called *structuples*) are items and are expected to capture specific levels of the construct. Thus, facet design not only provides a systematic way of defining the construct as a hierarchical continuum, but also offers an explicit connection between an item and its location along the continuum. Compared with the conventional instrument development process discussed in the earlier section, I argue that Guttman's facet theory as a methodological approach to theoretical thinking can offer a more systematic and efficient approach to capturing a construct that "leads to considerable conceptual clarity and control"



(Borg & Shye, 1995, p. x). It is more systematic and efficient in that there is an explicit and empirical connection between the construct definition and the manifestation of items (Borg & Shye, 1995; Guttman & Greenbaum, 1987).

In a study by Randal and Engelhard (2010), the researchers used facet theory and the sentence mapping technique to develop a 54-item instrument to measure teachers' grading philosophies - a unidimensional construct comprising four facets that influence teachers' grading philosophies: effort, ability, achievement, and behavior. Each item was written as a scenario that describes a student's situation in terms of the four facets and participating teachers were asked to give a letter grade and a numerical grade. A priori hypotheses on the locations of the items along the hierarchical continuum were made. Rather than using SSA, the many-facet Rasch model (Linacre, 1994) was used instead to check the psychometric properties of the items and to evaluate the instrument. The many-facet Rasch measurement model (MFRM) developed by Linacre (1994) is the extension of the Rasch dichotomous model. In the basic Rasch dichotomous model, the probability of a person giving a right answer (i.e., scoring 1 rather than 0) to an item is governed by the person's ability and the item's difficulty. The many-facet Rasch model includes one more parameter – raters' leniency in giving a rating to a person on a given item (Linacre, 1994). The purpose of the many-facets Rasch model is to adjust for systematic measurement errors that are contributed by factors occurred in measurement situation by introducing one (or more) component into the standard Rasch model (Linacre, 1994; Wolfe & Smith, 2007a). Accordingly, the model is rather robust to many forms of misfit. Since the purpose of this study is to build a reliable, valid, and meaningful scale to measure a unidimensional construct, the many-facet Rasch model is not suitable for this study.

Another study, discussed earlier as an application of Rasch measurement theory, by Ludlow and colleagues (2014) also applied facet design and the sentence mapping technique to develop a scenario-like scale to measure engagement in later life activities as a unidimensional construct. In this study, the construct of engagement consists of four facets (i.e., interest, focus, energy, and perseverance) and each contains three elements or levels (i.e., low, moderate, and high). A sentence mapping structure including four facets was used to construct scenario-like items. Unlike most studies applying facet theory that construct items representing all *structuples*, the researchers decided to only construct extreme scenarios to avoid cases of ambiguity (Borg & Shye, 1995; Ludlow et al., 2014). As a result, a total of nine scenarios were constructed: three from the higher level, three from the moderate level, and three from the lower level. According to Ludlow and colleagues (2014), the intention was to capture the boundaries in the levels of engagement at the initial stage of instrument development, and later to move on to capturing the subtleties between the extreme scenarios. Participants were asked to read each scenario-style item and assess their own engagement level against the scenarios by rating a five-point Likert scale (1 = *much less engaged than 'X'*, 2 = *less engaged than 'X'*, 3 = *about as engaged as 'X'*, 4 = *more engaged than 'X'*, and 5 = *much more engaged than 'X'*). The hypothesis is that scenarios representing the higher-level of engagement are harder to be 'endorsed' by the participants (i.e., rated as 'More engaged than X' or above) than scenarios representing the lower-level of engagement. The Rasch rating scale model was used as the analytical model to evaluate the instrument and check whether the a priori theory regarding the item locations was confirmed by the empirical data.

Applying an approach similar to Ludlow and colleagues (2014), this study utilizes the combination of Rasch measurement theory and Guttman's facet theory with the sentence mapping technique to develop scenario-style items to measure teachers' enactment of practice for equity. While the scenario-style items have the potential to capture the complexity and richness of teachers' reported practice for equity compared to the commonly see Likert-type items, it is not without challenges. Ludlow and colleagues (2014) have discussed the potential issues, including the length of scenarios that might cause respondents' fatigue as well as double-barreled items due to the multi-faceted nature. In the next chapter, I describe the steps taken to minimize these issues in greater details.

Following the Rasch principles, teachers' perceptions of their enactment of practice for equity is a unidimensional construct conceptualized as a hierarchical continuum of low to high-level of enactment. As discussed in the first section of this chapter, this construct consists of six interconnected principles of practice for equity, which are six main characteristics of the construct and are framed as six facets hereafter. Within each of the six facets, there are three elements or levels (i.e., low, moderate, and high). Given that this is the first stage of scenario item construction, I also choose to avoid cases of ambiguity (Borg & Shye, 1995, p. 40; Ludlow et al., 2014) and to focus on creating a "proof of concept" by capturing the extreme situations in enacting practice for equity (high and low) before attempting to capture the subtleties between the extremes. The sentence mapping technique is used to construct each of the scenario-style items that consist of multiple facets. Participants are instructed to reflect on their own practice, compare it against the practice of the individual teacher named in each scenario, and rate their own practice based

on a five-point Likert scale. The Rasch rating scale model is chosen as the statistical model of this study.

### **Summary**

Thus far, I have discussed some of the challenges encountered in research on teacher education in the first chapter. I also reviewed the relevant bodies of literature that justify the need for and the potential of a Rasch-based scenario-style instrument that measures teachers' self-reports about their enactment of practice for equity. Given the complex nature of the construct and intended purpose of the instrument, the research question of this study is: Can the construct of teachers' enactment of practice for equity be measured reliably and meaningfully by applying a novel approach to construct a Rasch-based scenario-style scale? In the following methodology section, I discuss the overall research design, the characteristics of target participants and the sampling approach, detailed procedure of how Rasch principles and facet theory guide the development of instrument, and the statistical model used.

## **CHAPTER THREE: METHODOLOGY**

In this chapter, I discuss the overall research design and the scope of work of this instrument development study, followed by a description regarding the target participants and the sampling approach used to recruit participants. Given the novel approach of this study, a section on procedure is dedicated to describing the detailed and lengthy process of developing the scenario-style items. The discussion on procedure includes the iterative process of construct clarification and a pre-pilot of 15 scenario-style items with feedback sessions. Next, I discuss the kind of information collected through the single self-report survey by laying out the plan of the survey, item types used and the purposes. Lastly, I discuss the statistical model chosen to analyze the data and the kind of psychometric results reported to address the research question.

### **Research Design**

This dissertation was intended to develop an instrument to measure teachers' self-reports about the extent to which they enact equity-centered practice in the classroom, included three phases. The first phase involved the development of scenario-style items and a pre-pilot study; the second phase involved pilot testing, and the final phase was the full-scale administration of the survey.

Phase one included the lengthy iterative process of construct clarification, a pre-pilot of 15 scenario-style items with key informants from both New Zealand and the U.S., two focus-group feedback sessions with the same group of key informants in the U.S., and written feedback via emails from informants who could not attend. This work is thoroughly discussed in the Procedure section, below. Phase two involved administering the pilot survey. This stage also included conducting the think-aloud exercise with three

key informants to understand any issues with poorly fitting items and to obtain ideas to better address issues that arose. In Phase three, a revised survey was administered to target participants in the U.S. Discussion regarding target participants, sampling procedure, data collection, and data analysis are most relevant to the pilot and the full-scale phases.

### **Participants**

The primary participants in this study were pre- and in-service teachers. Pre-service teachers who were included were either completing their practicum or in their last year of preparation programs. In-service teachers who participated had varying years of teaching experience in New Zealand and the United States. Given that 85% of teachers in the U.S. are prepared in programs located within higher education institutions (U.S. Department of Education, 2016), pre-service teachers from the U.S. were recruited from university-based undergraduate and graduate elementary or secondary programs and were being prepared to teach academic subject content areas. Similarly the in-service teacher participants were teaching academic subject content areas in K-12 school settings. Participants prepared through alternative routes, such as new educator preparation program providers like Boston Teacher Residency or alternative certification routes such as Teach for America, were excluded. The criteria also excluded special education, physical education, and art education teachers but included those who worked with English-Language Learners (ELLs).

### **Sampling**

The ideal sample size must be large enough to provide reliable and accurate estimates of parameters. According to Crocker and Algina (1986), a sample size of 200

is the minimum for an item analysis study. Others suggest that a general rule is to have five to ten times as many subjects as items (Hair, Black, Babin, & Anderson, 2010; Nunnally, 1967). Based on the suggested guidelines above, a sample size ranging between 75 and 150 participants was sought for both the pilot and the final full-scale administration. As detailed below, participants were identified and contacted through teacher preparation programs they were either enrolled in or had graduated from.

### **Pilot study**

The pilot study was conducted over the course of two weeks, between October 31 and November 14, 2016. A convenience sample was used for the pilot study, which included 360 pre-service teachers who were enrolled in teacher preparation programs and 600 novice teachers who recently (i.e., up to three years) graduated from the same programs at the University of Auckland. Participants were recruited through emails distributed by central communication services at the University of Auckland. A personal reminder was sent by a course instructor of the teacher preparation program during the two-week period.

In addition, I used the snowball sampling technique to recruit participants who were either pre-service or in-service teachers in the U.S. Specifically, I asked 29 teacher and teacher education colleagues to distribute a standard recruitment email to their colleagues and/or students. The recruitment message was also posted on relevant social media sites, including the Facebook page of the Teaching and Teacher Education Division of the American Educational Research Association, and the LinkedIn group of the Educational Research, Measurement, and Evaluation department. These participants

were not affiliated with Boston College – that is, they were not currently enrolled in or had graduated from a teacher preparation program at Boston College.

With regard to the think-aloud exercise, three key informants were identified based on their availability within a short period of time, their teaching experience and/or knowledge about the scenario-style items. One key informant was an ELL teacher in a public school, and the other two key informants were former high school English and Mathematics teachers. The three interviews were conducted between November 15 and 23, 2016. The key informants were first asked to respond to a select number of scenarios that included both good and poorly fitting items (i.e. statistically misfitting the psychometric model). Next, they were asked to explain their thought process while responding to the misfit items sentence-by-sentence, and were encouraged to point to potential issues in the misfit items. The key informants were also asked to provide feedback about how to better address the specific issues identified.

### **Final full-scale administration**

The full-scale administration was conducted between December 5 and December 20, 2016. A convenience sample was used for the final full-scale administration. The participants included 60 undergraduate and graduate-level teacher candidates who were enrolled in the secondary or elementary programs at Boston College and were in the full-practicum in the fall semester of 2016. These teacher candidates were recruited through multiple emails (i.e., one recruitment email and two reminder emails), an internal newsletter, and personal communication from the practicum director. The other pool of participants came from 149 novice teachers who graduated from teacher preparation programs at Boston College during the previous three years, and were recruited through a



recruitment email and three reminder emails sent by the researcher. In addition, ten teacher educators from other higher education institutions who were alumni of the Curriculum and Instruction doctoral program at Boston College agreed to help recruit their teacher candidates or program graduates that met the criteria. Using a similar recruitment approach, multiple emails were sent by these teacher educators to recruit and remind their students or former students during the two-week data collection period.

### **Procedure**

#### **Step one: Reviewing the syntheses and clarifying the construct**

The first step in developing valid and reliable items for any Rasch-based instrument is to have an in-depth understanding of the construct to be measured. To do so, I reviewed the five selected syntheses and programs of research that provided the basis of the development of the six facets of practice for equity (Grudnoff et al., 2017). While reviewing each of the five frameworks, I documented and summarized on separate excel sheets the pedagogical practice associated with positive student learning outcomes according to the organization structure used by each framework. The six facets identified earlier by the RITE research team members guided the review and construct mapping process. I categorized the description of pedagogical practice derived from each framework/program of research to one or more than one of the six preliminary facets for the initial construct mapping process. The “linking” between the six facets and the pedagogical practice from all six syntheses/programs of research was then completed on a separate excel sheet. This complete excel sheet presents and summarizes the pedagogical practice collected from across all five syntheses and programs of research in rows for each facet in columns (see Appendix A on page 184 for the illustration of part of

the excel sheet). Examining the practice categorized under each column of facet (i.e., across rows for each column) provides a primitive collection of teaching practice for each facet (see Appendix B on page 198 for the collection of practice for all six facets).

This process of construct mapping was discussed and confirmed with members of the project RITE research team, who are experts in teacher education. During the initial construct mapping process, several questions and concerns emerged, and were discussed and resolved within the research team. For example, during the mapping process, we recognized that a pedagogical practice identified from a synthesis could sometimes be categorized under more than one facet. This was not surprising, given that the six facets are interrelated to each other, in both a conceptual and practical sense. The team made the decision to categorize a pedagogical practice under only the one most relevant facet for the purpose of instrument development. Another concern raised was that Facet 4 - taking an inquiry stance through using evidence to scaffold learning and improve teaching – seemed to be an overarching facet of the first three, which created some difficulty in the process of categorization. The research team decided to conceptualize and rename Facet 4 “assessing for learning and using evidence to assist teaching and to scaffold student learning.”

A third major concern was that it was more difficult to identify pedagogical practice associated with Facet 5 and Facet 6. For Facet 5, (i.e., taking agency/responsibility for further professional engagement and learning) this could be because most existing instruments and protocols were developed to assess teachers’ practice in the classroom. Unlike the teaching practice for the first four facets that could be identified almost across all five syntheses, I specifically reviewed the Teaching and

Learning Research Project (James & Pollard, 2006) to identify relevant practice for Facet 5. With advice from the panel of content experts, I also revised the name of Facet 5 to “taking an inquiry stance for further professional engagement and learning.”

A similar challenge was encountered when I reviewed the syntheses and programs of research to identify relevant practice associated with Facet 6. With advice from the context experts, I looked into Te Kotahitanga Effective Teaching Profile (Bishop, Berryman, Cavanagh, & Teddy, 2009) to identify relevant practice associated with Facet 6. It is worth noting that teaching practice which recognizes and challenges inequities is absent in Danielson’s Framework for Teaching for Classroom Observation used in the Measurements of Effective Teacher in the U.S. (MET Project, 2013) and the Center for Research on Education, Diversity, and Excellence’s [CREDE] five standards for effective pedagogy (Dalton, 2007). In addition, following advice from the content experts, I revised the name of Facet 6 to “recognizing and challenging inequity” as there exist many factors (e.g., pressure from accountability policy) that can prevent a teacher from challenging inequity.

Discussion with the research team and review with the content experts occurred in parallel with the construct mapping process. Once the initial construct mapping was completed with all pedagogical practice identified from all five syntheses and programs of research categorized under one single facet, I conducted one final cross-check to identify the discrepancies between my categorization and how the content experts “coded” the teaching practice when they formulated the six themes. At this point, the discrepancies were fairly minor, which indicated a desirable level of inter-rater reliability. I was able to work with two content experts on the research team to resolve the remaining

discrepancies easily and without serious disagreement. The discussions among research members and the decisions made during the construct mapping process were critical to ensure that the facets were clear, had good coder reliability, and were related yet distinct from each other as suggested by facet theory (Borg & Shye, 1995; Guttman & Greenbaum, 1998). As a result of the iterative construct mapping process, I had six clearly defined facets and a collection of pedagogical practice that belonged to each facet, as well as a clear definition for teaching practice that recognized and challenged inequities as a stand-alone facet (i.e., Facet 6). The intention here was to address the often diluted or arbitrarily defined political aspect of teaching practice commonly seen in the existing instruments.

**Step two: Applying a systematic approach to developing rich narratives and descriptions on three levels of enactment for each facet**

With the construct map completed, I developed a general description in a rich narrative form for each facet, as well as descriptions for three levels of teaching equity enactment (i.e., low, moderate, and high) for each facet. To develop the general description, I read the teaching practice categorized under each facet multiple times, and highlighted the similarities/recurring themes mentioned by different syntheses under each facet. These themes became the main characteristics within each facet, and were used as a placeholder for a group of pedagogical practice when I developed the general description ranging between one to two pages for each facet. The first version of general description is in bullet-point form, where pedagogical practice was categorized into specific characteristics. The sources where the practice was derived from were also

documented (see Appendix C on page 224 for the first version of the general description of all six facets).

Using the first version of the general description for each facet, I developed the second version in a page-long rich narrative form (see Appendix D on page 232 for the rich narratives of all six facets). Based on these rich narratives, I then developed descriptions to capture the three levels of enactment (i.e., *low* – level 1, *moderate* – level 2, and *high* – level 3) for each facet (a description of levels of enactment for all six facets can be found in Appendix E on page 238). This step was guided by the Rasch measurement principles that conceptualize a unidimensional variable, such as teachers' enactment of practice for equity in the form of order – lower levels of enactment to higher levels of enactment (Wright & Masters, 1982). To develop scenarios (each becomes an item itself as will be shown later on) that measure the variation (or levels) of teachers' enactment of practice for equity, I needed to have descriptions that captured the different levels within each facet. I focused on three distinct levels (i.e., low, moderate, and high) because I intended to capture the boundaries of the construct given the initial stage of instrument development and novelty of this approach before capturing the subtleties. These three levels within each facet are elements or *structs* in the language of facet theory (Borg & Shye, 1995). The same strategy was also used in a previous study that successfully developed scenario-style items to measure engagement in later-life activities using Rasch and facet theory (Ludlow et al., 2014). The documents produced in this step were critical in that they provided the stepping stones for creating scenarios which eventually became scenario-style items, as I will explain shortly.

### **Step three: Determining the scale's structure for scenario development**

The decisions on the scale's structure for scenario development involved two main issues. First, with three enactment levels and six facets, there exist 729 possible combinations of facets (or *structuples*). The problem was whether it is feasible and reasonable to capture all 729 possible combinations at the beginning stage of the instrument development. I decided to focus on extreme group contrasts, meaning that I constructed extreme scenarios from the highest, moderate, and then lowest levels of facets. The goal here was to avoid cases of ambiguity (Borg & Shye, 1995, p. 40; Ludlow et al., 2014) and to create the *proof of concept* by capturing the extreme situations in enacting practice for equity before attempting to capture subtleties between the extremes.

The second issue was that given the complexity of the construct and the length of a scenario, it was impossible to include all six facets in each scenario as they would be too long. I decided that including three facets in each scenario was sufficient and reasonable. To select the combinations of three facets systematically, I decided that Facet 6 (i.e., recognizing and challenging inequities) would appear in all combinations. This decision was informed by the theoretical lens of this study, which puts equity front and center. That is, practice that recognizes and challenges inequities must be present in equity-centered teaching practice and thus must appear in all scenarios. The other two facets (besides Facet 6) in any scenario were chosen in such a way that each scenario had one overlapping facet with another scenario. Therefore, the selection of facets (particularly Facets 1 through 5) appears like a spiral. Table 3.1 below presents the structure of scenario development.

Table 3.1

*Scenario Development Structure*

Level of Enactment	Scenario #	Facet 1	Facet 2	Facet 3	Facet 4	Facet 5	Facet 6	Total
High Level of Enactment (Level 3)	Scenario 1	X (L=3)	X (L=3)				X (L=3)	9
	Scenario 4		X (L=3)	X (L=3)			X (L=3)	9
	Scenario 7			X (L=3)	X (L=3)		X (L=3)	9
	Scenario 10				X (L=3)	X (L=3)	X (L=3)	9
	Scenario 13	X (L=3)				X (L=3)	X (L=3)	9
Moderate Level of Enactment (Level 2)	Scenario 2	X (L=2)	X (L=2)				X (L=2)	6
	Scenario 5		X (L=2)	X (L=2)			X (L=2)	6
	Scenario 8			X (L=2)	X (L=2)		X (L=2)	6
	Scenario 11				X (L=2)	X (L=2)	X (L=2)	6
	Scenario 14	X (L=2)				X (L=2)	X (L=2)	6
Low Level of Enactment (Level 1)	Scenario 3	X (L=1)	X (L=1)				X (L=1)	3
	Scenario 6		X (L=1)	X (L=1)			X (L=1)	3
	Scenario 9			X (L=1)	X (L=1)		X (L=1)	3
	Scenario 12				X (L=1)	X (L=1)	X (L=1)	3
	Scenario 15	X (L=1)				X (L=1)	X (L=1)	3

\*Note: X indicates that the specific facet is included in the scenario

For instance, for the high level of enactment (i.e., level 3), Scenario 1 includes Facets 1, 2 and 6, and Scenario 4, picking up from Facet 2, includes Facets 2, 3, and 6. This approach created a chain of linkage and allowed the scenarios to cover the domain of the construct without overwhelming the respondents. This was done for all three

levels of teachers' enactment of practice for equity. As a result, there were a total of 15 scenarios with five scenarios capturing each of the three distinct levels of enactment of practice for equity. Scenarios 1, 4, 7, 10, and 13 capture the high-level of enactment; scenarios 2, 5, 8, 11, and 14 capture the moderate-level of enactment; and scenarios 3, 6, 9, 12, and 15 capture the low-level of enactment.

As shown in Table 3.1, the ordinal codes assigned to the three enactment levels (i.e., *low* = 1, *moderate* = 2, and *high* = 3) across the six facets indicate the overall enactment level of each scenario. The column on the far right hand side is the sum of the coded level for each scenario. For example, the five scenarios (Scenarios 1, 4, 7, 11, and 14) capturing the high-level enactment each have a sum score of nine (i.e., each scenario consists of three facets written at level-3). This means that these five scenarios are hypothesized to be the hardest for participants to indicate that they enact at the same or higher level of practice for equity than the practice described in these five scenarios (the specific response options are discussed in step four below). Similarly, the five scenarios (Scenarios 3, 6, 9, 12, and 15) capturing low-level enactment each have a sum score of three. This indicates that these five scenarios are expected to be the easiest for participants to indicate that they enact at the same or higher level of practice than the practice described in these five scenarios.

Therefore, I hypothesized that the 15 scenarios would form a three-cluster structure with the five high-level scenarios clustering on the top of the construct's hierarchical continuum presented in the variable maps, followed by the five moderate-level scenarios clustering in the middle and the five low-level scenarios at the bottom. Since each scenario in each enactment level only covered three out of six facets, to



completely cover the domain of the construct, it was reasonable to have five scenarios capturing each of the three distinct levels of enactment. This *a priori* hypothesis on scenario structure and spread was used for the first stage of instrument development.

#### **Step four: Formulating sentence mapping structures and constructing scenarios**

Once the structure of scenario development was determined, I developed *structs* for each facet, meaning phrases or sentences that represent ranges (i.e., *low*, *moderate*, and *high*) within each facet. I used the previously developed descriptions of three levels of practice and shortened the descriptions for each characteristic under each facet. Again, because of the complexity of the construct, I decided to break the description for some characteristics into two parts (which I called *sub-characteristics*). Like the work engagement scenarios (Ludlow et al., 2014), I also chose to use *unobstructive facetization* which avoided using common range terms such as low, moderate, and high but instead used words that convey the similar meaning. The *unobstructive facetization* can also make the scenarios more conversational and meaningful to the respondents.

Once the phrases or sentences to represent *structs* of each facet were created, the goal of the sentence mapping technique was to assemble *structs* by using the formal and informal components. Given that there were five combinations of scenarios, a different sentence mapping structure was used for each of the five combinations. However, one sentence mapping structure was used to develop low, moderate, and high-level enactment scenarios of the same combination (see Appendix F on page 255 for the sentence mapping structure of each combination and the 15 scenarios).

Below is an example of the sentence mapping structure used for the combination of Facets 1, 2, and 6. To construct a scenario, the bolded parts in the parentheses were filled in with descriptions of practice drawn from selected facets. A person's name for each scenario was used to make the scenario authentic. The parts that are not bolded stayed constant and served to connect the bolded parts and to provide a context. Using the sentence mapping structure example, the three scenarios capturing the high, moderate, and low levels of enactment of Facets 1, 2, and 6 are as follows.

Example. Sentence Mapping Structure

(**Person**) holds (**Facet 1, Characteristic 1**). He/she sees the home culture brought by students (**Facet 2, Characteristic 1.1**), and (**Facet 2, Characteristic 1.2**). (**Person**) (**Facet 6, Characteristic 1.3**). In the classroom, (**Person**) (**Facet 2, Characteristic 2**) and (**Face 1, Characteristic 2**). He/she is (**Facet 1, Characteristic 3**).

Scenario 1 capturing the high-level of enactment of Facets 1, 2, and 6:

**Maria holds high expectations for all her students and communicates challenging and meaningful learning goals to all students clearly. She sees the home culture brought by students as the strength and assets for their learning, and collaborates closely with parents/caregivers and community members as the partners of students' learning. Maria encourages students to take initiatives on the content, assessment, and directions of learning. In the classroom, Maria consistently draws students' prior knowledge and culture and purposefully designs relevant and valuable learning experiences for all students. She is**

**skillful in pulling together a range of instructional approaches, techniques, strategies, and moves with clear and error-free explanation to stimulate students' learning, interests, and motivations.**

Scenario 2 capturing the moderate-level of enactment of Facets 1, 2, and 6:

**Tim holds high expectations for some students, and generally communicates the learning goals clearly.** He sees the home culture brought by students **as the strength but sometimes engages in stereotypical thinking,** and **collaborates closely with some parents/caregivers as the partners of students' learning.** **Tim sometimes encourages students to take initiatives on the content, assessment, and directions of learning.** In the classroom, **Tim occasionally embeds students' culture into learning experiences which can be limited to motivate learning and the design of learning experiences is sometimes not deliberate and irrelevant to students.** His **content knowledge is inconsistent with some errors and lack of clarity, and explanation of concepts is not always compelling to capture students' interest.**

Scenario 3 capturing the low-level of enactment of Facets 1, 2, and 6:

**Kevin holds high expectations for only a few students and communicates the learning goals ambiguously.** He sees the home culture brought by students **as a problem and weakness for their learning,** and **does not include parents/caregivers and community members as the partners of students' learning.** **Kevin determines the content,**

**assessment, and directions of students learning with limited student involvement. In the classroom, Kevin occasionally uses students' prior knowledge and culture and rarely designs relevant and stimulating learning experiences for all students. He utilizes a limited range of instructional approaches and techniques; the explanation of key concepts contains some errors and does not capture students' interest.**

As developed, this is a self-report type of instrument. Figure 3.1 below presents what a scenario-style item is like using the high-level of enactment for Facets 1, 2, and 6 as an example (For brevity reasons, the terms 'scenarios' or 'items' will be used instead of 'scenario-style items' hereafter). Above each scenario-style item, a question guides participants to reflect on their own practice and compare it against the example teacher's practice in the scenario. Participants are asked to rate based on a five-point Likert scale (*much lower* = 1, *slightly lower* = 2, *about the same* = 3, *slightly higher* = 4, *much higher* = 5). Choosing *about the same* means that participants are aligned in their practice to the practice of a specific scenario; choosing *slightly lower* or *much lower* means that participants consider their practice to be at a lower enactment level than the scenario; and choosing *slightly higher* or *much higher* means that participants consider their practice to be at a higher enactment level than the scenario. Higher scores indicate higher levels of enactment of equity-centered practice. It was intended that scenarios representing higher levels of enactment would be harder for respondents to reply with higher or much higher ratings than scenarios of lower-level enactment.

Figure 3.1

*An Example of Scenario*

**Scenario-style Item**

What is your assessment of your own level of enactment compared to that teacher's practice described in that scenario?

Maria holds high expectations for all students and clearly communicates challenging and meaningful learning goals. She sees students' home culture as an asset and collaborates closely with parents/caregivers as partners. Maria encourages students to take initiative regarding their learning. In the classroom, she consistently draws on students' prior knowledge and culture and purposefully designs relevant learning experiences for all. Maria skillfully uses a variety of instructional approaches to motivate students' learning. Her explanations are clear, compelling, and accurate.

**Much lower   Slightly lower   About the same   Slightly higher   Much higher**

**Step five: Conducting a pre-pilot and feedback sessions**

For the first draft of scenarios, the length of scenarios had word counts ranging between 95 and 125. To reduce the length of scenarios and minimize the burden on survey respondents, I revised the 15 scenarios with suggestions from the content experts. Each of the revised scenarios had a reduced word count of approximately 80 words and still followed its own sentence mapping structure. I then conducted a pre-pilot of the 15-item instrument on Qualtrics with a group of key informants - teacher educators and doctoral students with content and/or measurement expertise ( $N = 16$ ) in the Lynch School of Education at Boston College, as well as one teacher candidate enrolled in a preparation program at the University of Auckland in New Zealand. The specific purpose of working with a key informant from New Zealand was to ensure that the use of language in instructions, items (both scenarios and demographic questions), and response options would be appropriate and make sense in the New Zealand context. I also held two one-hour feedback sessions with eight of the pre-pilot participants to seek their

reactions to the instrument (i.e., the instruction, the scenarios, and the response options) and suggestions (see Appendix G on page 265 for the guiding questions for written or in-person feedback sessions). I received written feedback from the rest of the informants who participated in the pre-pilot but were unable to attend the feedback sessions.

Two major issues arose based on informants' feedback. First, some descriptions of practice were too general and abstract which became repetitive and less engaging for the participants (i.e., teacher candidates with classroom teaching experiences and in-service teachers). For example, descriptions such as *setting high expectations* or *encourage students to take initiatives on their own learning* are vague. Another related issue was that the sentences in a scenario were not integrated well enough to read like a story, which created the problem of multi-barreled meanings. The second issue was the effect of social desirability, especially for low-level scenarios where negative wordings were used to describe the low-level enactment of practice for equity. The resulting problem was that a respondent might never choose *about the same* to negatively worded situations.

Participant feedback was used to refine the items for the pilot study (see Appendix H on page 267 for the comparison of the original and the revised scenario-style items for the pilot). To address the first issue, scenarios were revised to be more holistic, specific, and story-like while relying on the same sentence mapping structure as much as possible. Feedback from key informants suggested that making each scenario more engaging also would help reduce the burden on participants, even though the length of scenarios stayed more or less the same. To address the issue of social desirability, the negative wording used in the low-level scenarios was replaced with phrases that still

conveyed similar conditions but without the negative tone. Below is an example of both the original and revised version of a negative scenario. In addition, an existing validated scale measuring social desirability (Strahan & Gerbasi, 1972) was added to the pilot survey to investigate the potential influence of social desirability on participants' responses. The revised scenarios were then reviewed by the panel of content experts again to ensure that they conveyed the same meaning and levels of enactment.

Original version of Scenario 3 (low-level enactment of Facets 1, 2, and 6):

Kevin holds high expectations for very few students and does not communicate them clearly. He sees students' home culture as an obstacle and rarely engages with parents/caregivers as partners. Kevin hardly encourages students to take initiative regarding their learning. In the classroom, Kevin seldom draws on students' prior knowledge and culture to design learning experiences that are relevant and stimulating. He utilizes a very limited range of instructional approaches to explain key concepts. His explanations contain major errors and do not capture students' interests.

Revised version of Scenario 3:

Kevin sets achievable goals for his students but finds it hard to communicate them. He sees students' home culture as challenging and doesn't expect parents to be his partners in teaching. Kevin sets out lessons for his students so they know what they need to do. In the classroom, Kevin uses textbooks and self-designed learning experiences that he believes deliver the appropriate curriculum. He utilizes a few instructional

approaches to explain key concepts, which he feels unsure at times and struggles to capture students' interest.

In addition, I received general suggestions about the survey instructions and the survey layout to provide clearer guidelines and to make the entire responding experience more manageable for the participants. For instance, some recommended that including an instruction page with a practice item before the 15 scenarios could be useful to avoid the “Start-up” effect. A similar recommendation was made by a previous study using the same measurement strategy to capture engagement levels in later life activities by using scenarios (Ludlow et al., 2014). The “Start-up” effect exists when survey respondents are presented with unconventional item format, and there is some initial confusion about how to respond to the item. This confusion can be reflected in the opposite options (i.e., unexpectedly low or high responses) being chosen to the first few items. For example, survey participants might not be accustomed to the “comparative” nature of the task and choose “Much lower (1)” to a low-level scenario when they are expected to be “Much higher (5)” than the low-level scenario. Therefore, I included a brief survey instruction followed by a moderate-level practice item before the 15 scenarios.

Besides the suggestion of adding a moderate-level practice item, some informants proposed that starting the scenario-item survey section with a moderate-level scenario might also help minimize the start-up effect. The rationale was that scenarios capturing either the low or the high level of enactment might be more likely to prompt the participants' to react to the scenario and subsequently forget the survey instruction. For this reason, I included a moderate-level practice item and used a moderate-level scenario



as the first item. I also incorporated other general suggestions, such as adding a progress bar and retaining the layout of three scenarios in one page, into the pilot survey.

## **Data Collection**

### **Instrumentation**

In both Phase two and Phase three of the study, data were collected using a single self-report questionnaire on Qualtrics. Since the instrument employed a novel measurement approach to measure teachers’ self-reports about their enactment of practice for equity, a complex construct, the questionnaire included other items in addition to the instrument of 15 scenario-style items to address issues such as potential start-up effect and social desirability. The questionnaire comprised three major sections: 1) One practice scenario followed by 15 scenarios, 2) 20 true-false items assessing the level of social desirability, and 3) five multiple-choice items on demographics and teaching contexts. Table 3.2 below presents the section plan of the survey (see Appendix I on page 277 for the pilot survey and Appendix J on page 292 for the final survey). I discuss each section below – the kind of information collected, scales used, and the purposes – in detail following the sequence that the items were presented to the participants.

Table 3.2

#### *Survey Section Plan*

	<b>Section Scale</b>	<b>Number of Items</b>
<b>Section I</b>	Scenario-style items assessing teachers’ self-reports about their enactment of practice for equity	16
<b>Section II</b>	Social desirability scale	20
<b>Section III</b>	Participants’ demographic information and teaching background	5

***Scenario-style items.*** The first section of the questionnaire was the instrument developed in this study to measure teachers' self-reports about their enactment of practice for equity. There were a total of 15 scenarios with five items in each of the three levels of enactment of practice. The participants were given instructions to reflect on their own practice and compare it against each individual teacher's practice in a given scenario. Five-point Likert response options included *much lower* (1), *slightly lower* (2), *about the same* (3), *slightly higher* (4), and *much higher* (5). Higher scores indicated higher levels of enactment of practice for equity. Scenarios of higher-level enactment would be harder for the participants to give higher ratings than scenarios of lower-level enactment. In addition to the 15 scenario-style items, a practice item capturing the moderate level of enactment of practice for equity was added at the beginning of the questionnaire to address the potential start-up effect.

***Social desirability scale.*** The second section of the questionnaire was a validated and widely used scale measuring the level of social desirability (Strahan & Gerbasi, 1972). It was unavoidable for participants to perceive higher levels of enactment as being more desirable than the lower levels of enactment, even with efforts to minimize the effect of social desirability through avoiding negative wordings. Therefore, including an established scale of social desirability allowed an investigation into how social desirability might influence responses. This investigation offered information about the extent to which an irrelevant construct such as social desirability, as a source of systematic measurement error, might influence the way respondents answered an item. The 20-item social desirability scale by Strahan and Gerbasi (1972), a modified scale from the original 33-item scale developed by Crowne and Marlowe (1960), was used to

minimize the response burden on the participants. Both reliability and criterion-related validity were tested across four different samples and conditions of questionnaire administration (Strahan & Gerbasi, 1972). The 20-item social desirability scale has a Kuder-Richardson formula 20 (KR-20) reliability coefficient of approximate 0.8 (Strahan & Gerbasi, 1972). The correlation coefficients between the 20-item scale and the Crowne-Marlowe social desirability scale are about 0.9 across all four samples (Strahan & Gerbasi, 1972).

***Demographics and teaching background.*** The last section of the questionnaire included items about demographic information and participants' teaching background. A structured-response format was used for items where participants were asked to select their response from a list of options. Demographic items included gender, race, years of teaching experiences, level and subject matter of teaching (i.e., elementary and secondary).

### **Data Analysis**

As indicated earlier in Chapter One, the primary research question was: *Can the construct of teachers' self-reports about their enactment of practice for equity be measured reliably and meaningfully by using a Rasch-based scenario-style scale?*

The Rasch rating scale model was used to analyze data collected for the 15 scenarios (Andrich, 1978; Wright & Masters, 1982). The statistical model of the rating scale is presented below (See Equation 3.1). The 15 scenario-style Likert-scale items yielded total scores for each person and item. Based on the rating scale model, these raw scores were transformed into logits, which are the statistical estimates for 1) a respondent's level of enactment of practice for equity, 2) an average difficulty of enacting

each of the 15 items, and 3) the difficulty of giving responses from one category to the next, which are the threshold estimates. In the rating scale model, the “threshold parameter” is estimated the same for all items in a scale and the difficulty of moving from one step to another is in successive order – the second step is intended to be more difficult than the first, and the third step is intended to be more difficult than the second, and so on.

$$\pi_{nix} = \frac{e^{\sum_{j=0}^x n_i [\beta_n - (\delta_i + \tau_j)]}}{\sum_{k=0}^m e^{\sum_{j=0}^k [\beta_n - (\delta_i + \tau_j)]}} \quad (3.1)$$

Where

- $\pi_{nix}$  = probability of person n responding in category x to item i
- $\beta_n$  = enactment level for person n
- $\delta_i$  = enactment difficulty for item i
- $\tau$  = threshold parameter, or the location of kth step in each item relative to the item’s scale value
- $x = 0, 1, \dots, m$  for the category response options
- $e$  = a transcendental number whole value, rounded to three decimal places (2.718).

If the empirical data fit the Rasch model, a respondent enacting a higher level of practice for equity is more likely to respond *about the same* or higher to more items and will have a positive person logit; a person enacting lower level of practice for equity is less likely to respond ‘*higher*’ to most items and will have a negative person logit. On the other hand, for a more difficult item (i.e., capturing higher levels of equity-centered practice), individuals are more likely to score low on the item, which gives a positive

logit to the item. For an easy item (i.e., capturing lower levels of equity-centered practice), individuals have a higher chance to score high on the item and the item will have a negative logit. Given the five-point Likert scale used in the instrument, there are four steps for each item and the difficulty of moving from one step (e.g., *much lower*) to the next (e.g., *slightly lower*) is intended to be in order for all items. The WINSTEPS software package was used to conduct the analyses (Linacre, 2015, v3.91.2).

The Rasch rating scale model is a confirmatory test of the extent to which the hypothesized scenario structure is supported by the empirical data. It is expected that the five scenarios for each of the three distinct levels (i.e., high, moderate, and low) would form three clusters along the hierarchical continuum of the construct. Specifically, the five scenarios of low-level enactment should cluster at the lower end of the continuum, the five scenarios of moderate-level enactment should cluster in the middle, and then the five high-level ones should cluster at the upper end of the continuum.

Psychometric results included the “variable maps” which graphically present the order and spread of calibrated items based on the empirical data. The variable maps provide information regarding: a) whether the intended order (i.e., three-cluster structure) was confirmed by the calibrated items, b) whether items were sufficiently spread out to define the variation of the teachers’ enactment of practice for equity, c) whether respondents were well spread (or not – teacher candidates at either Boston College or the University of Auckland New Zealand might be rather homogeneous given the social justice and equity focus of the programs), d) whether the difficulty levels of calibrated items seemed to be on target with the ability (i.e., enactment level) of respondents, and e) what it meant to move from low-level enactment to high-level enactment? When

calibrated items fit the model and the *a priori* theory, the variable map can provide construct validity evidence of the scale.

Category Characteristic Curves (CCCs) are used to check whether the response categories functioned as intended (Wolfe & Smith, 2007b). Specifically, this figure provided information regarding: a) whether the five response categories were used by respondents and whether each was dominant in one area of the engagement level distribution for the participants, and b) whether the probability of moving from one category to the next was in the expected ordered pattern. Rasch-Andrich category estimates (Andrich, 1978), which indicate the 50% probability of moving from one category to the next, are reported.

Fit statistics for items and persons are presented as well (Wolfe & Smith, 2007b). As mentioned earlier, a more liberal threshold of 1.2 ~1.3 for both unweighted and weighted fit statistics together with the t statistics (i.e.,  $t > |\pm 2|$ ) or 3 (i.e.,  $t > |\pm 3|$  depending on the sample size) are used to identify misfit items. Further investigation into the misfit items, such as the item difficulty, entry order, or confusion on the instruction, was conducted to provide an explanation for the misfit and information for further revision. Likewise, the same misfit criterion was used to identify persons who gave unexpectedly high or low responses and possible explanations (e.g., individual's carelessness, a systematic pattern among a particular group of respondents) for the misfit. The separation index and reliability estimate for both items and persons are also reported.

I also conducted a residual analysis to check whether there was any non-random factor in the standardized residuals (Ludlow, 1986). I obtained the standardized residuals and performed Factor Analysis with Principal Axis Factoring method on the standardized

residuals in SPSS. I reported statistics including the size of the determinant, the Kaiser-Meyer-Olkin (KMO), the significance level of Barlett's test, factors with eigenvalues larger than one, the percent of variance explained by factors with eigenvalues greater than one, and the scree plot. The determinant is an indicator of whether there exists enough variability in a correlation or covariance matrix. A non-zero determinant means that the matrix has sufficient variability for factory analysis. In the case of residual analysis, it is less desirable to see a non-zero determinant. Kaiser-Meyer-Olkin (KMO) is a measure of sampling adequacy. When a KMO approaches 1, this suggests that items have small partial correlations with the common factor and items share a common factor. When a KMO approaches 0, this suggests that the partial correlations between items and the factor increase and items do not share a common factor. A KMO of larger than 0.7 is considered sampling adequate; however, in the case of residual analysis, it is less desirable to see a large KMO. Barlett's Test of Sphericity examines whether there is significant inter-item correlation variation in the item correlation matrix. For the purpose of residual analysis, it is preferable to see a non-significant Barlett's test. The scree plot provides a visual representation of the underlying factor structure. For the purpose of residual analysis, ideally, there should not be an apparent elbow shape or any factors with eigenvalue larger than one extracted.

## CHAPTER FOUR: RESULTS

This chapter presents the results of analyses for the pilot study and the final full-scale administration, including an overview of participants' responses (e.g., response rate, how missing data were addressed), descriptive statistics, and the results of Rasch analyses. The discussion also includes detailed steps taken to revise and refine the scenario scale. The research question for this study was: *Can the construct of teachers' self-reports about their enactment of practice for equity be measured reliably and meaningfully by using a Rasch-based scenario-style scale?*

### Pilot Study

#### Overview of participants' responses

As previously outlined in Chapter Three, the pool of participants for the pilot study included both current students and recent graduates of the teacher preparation programs at the University of Auckland in New Zealand. A recruitment email was drafted by the author and disseminated by central communication services at the University of Auckland to a total 960 pre- and in-service teachers, and 47 people (4.9%) responded to the pilot survey. Since the two-week data collection period occurred during the final weeks of the semester at the University of Auckland, the low response rate was not surprising. A second pool of participants for the pilot study was recruited through a snowball sampling approach. They were pre-service and in-service teachers in the U.S. who were not current students or graduates of a teacher preparation program at Boston College. A total of 74 pre- and in-service teachers responded to the survey. Altogether, 121 participants responded to the pilot survey.



**Missing data.** Among the 121 participants who responded to the survey, 67 (55.4%) responded to all 16 scenarios (one practice item and 15 scenarios) capturing teachers' self-reports about their enactment of practice for equity in the classroom. Six participants (5%) skipped one of the 16 scenarios and 48 participants (38.8%) did not respond to at least three items. A further investigation into the 48 participants who skipped at least three scenarios showed that 19 respondents (15.7% of 121) agreed to participate in the survey but did not complete any items following consent; 16 respondents (13.2% of 121) completed the practice item only; and, ten respondents (8.2% of 121) completed the practice item and only the first few scenarios. See Table 4.1, below, for an overview of missing data.

Table 4.1

*Number Missing: Pilot Study*

<b>Number of Missing Variables</b>	<b>Frequency (Number of Cases)</b>	<b>Percent</b>	<b>Cumulative Percent</b>
0	67	55.4	55.4
1	6	5.0	60.3
3	3	2.5	62.8
6	4	3.3	66.1
9	2	1.7	67.8
12	4	3.3	71.1
15	16	13.2	84.3
16	19	15.7	100.0
<b>Total</b>	121	100.0	

Examining the number of dropouts and number of items missing provided plausible explanations to the dropout pattern. It is likely that the dropout participants were overwhelmed by the amount of effort required to respond to the scenarios after they gave consent and read the survey instructions and practice item. The information presented in Table 4.1 supports this rationale given that 19 respondents consented to

participate, but left the survey without answering the practice item, and an additional 16 participants left after answering the practice item. For those who did continue, it is possible that the amount of effort required to answer the first six items on the first two pages of the Qualtrics survey caused abandonment of the remainder of the survey. Of the 48 participants who skipped at least three items, none persisted to the last scenario item and continued to the rest of the survey, which included a 20-item social desirability scale and five additional questions about demographic information and teaching background. The six participants who skipped only one scenario completed the rest of the survey. That is, all 73 participants (67 with complete responses, 6 with one missing item) that made it through the scenario section also completed the rest of the survey.

To further examine whether data were missing systematically, I checked the number of missing values based on country (i.e., New Zealand or the U.S.). As shown in Table 4.2, among the 47 participants from New Zealand, 24 (51.0%) did not complete the survey; 23 participants completed the survey, including two participants who skipped one scenario. Among the 74 U.S. participants, 24 (32.4%) participants did not complete the survey, and 50 (67.5%) completed the survey, including four participants who skipped one item. The different response rates might be explained by the timing of data collection (i.e., the end of semester for New Zealand participants), and the recruitment processes (i.e., only one recruitment email was sent to the New Zealand participants while multiple emails and personal communication were used to recruit the U.S. participants).

Table 4.2

*Number Missing by Country: Pilot Study*

<b>Number of Items Missing</b>	<b>United States</b>	<b>New Zealand</b>	<b>Total</b>
0	46	21	67
1	4	2	6
3	3	0	3
6	1	3	4
9	1	1	2
12	0	4	4
15	8	8	16
16	11	8	19
Total	74	47	121

I also examined whether there was a systematic pattern between the number of dropouts and their responses to the practice item, considering that 16 participants (13.2% of 121) dropped out after answering the practice item. Tables 4.3 presents the number of missing values per each of the response categories for the practice item. As shown in Table 4.3, more participants who chose *about the same* ended up dropping out of the survey later on; however, those that completed the survey also selected that answer. It is still possible that those dropout participants found the scenarios too overwhelming and/or were unsure how to respond given the multiple-sentence nature of the items.

Table 4.3

*Number Missing by Response Categories of the Practice Item: Pilot Study*

Number Missing	Response Category					Total
	Much lower	Slightly lower	About the same	Slightly higher	Much higher	
<b>0</b>	0	8	26	23	10	67
<b>1</b>	0	0	2	3	1	6
<b>3</b>	0	2	0	1	0	3
<b>6</b>	0	0	2	2	0	4
<b>9</b>	0	0	1	0	1	2
<b>12</b>	1	1	0	1	1	4
<b>15</b>	0	4	7	3	2	16
<b>Total</b>	1	15	38	33	15	102

Based on the results of examining the missing data, I decided to use the *Listwise Deletion method* but retained the six participants who only skipped one scenario item. I deleted cases with more than one missing value because these participants did not continue responding to the rest of the pilot survey and I did not have demographic data, teaching background, and their social desirability scores. This decision produced a 60.3% (73 out of 121) survey completion rate and led to a total of 73 participants in the item analysis for the pilot.

## **Descriptive statistics**

*About the participants.* Among the 73 participants in the analysis, 50 were pre- or in-service teachers in the U.S. and 23 were pre- or in-service teachers in New Zealand. The majority of participants were female (82.0% for U.S. and 86.9% for New Zealand). Most participants were White (96.0% of the U.S. participants) or European (65.0% of New Zealand participants). Among the 23 New Zealand participants, six participants (26.0%) identified themselves as Asian and one (4.3%) identified themselves as Pacific Islander.

Most New Zealand participants were pre-service teachers with less than one year of teaching experience (65.2%). Seven participants had one to three years of teaching experience (30.4%), and one participant had five to ten years of teaching experience. Among the 50 participants in the U.S., 24 (48.0%) had more than ten years of teaching experience, 19 (38%) had five to ten years of teaching experience, five (10%) had one to three years of teaching experience, and two (4%) had less than one year of teaching experience. Given the sampling approaches I used and the sources of participants in the two different contexts, the different composition of participants in terms of their classroom teaching experience was not surprising.

The majority of New Zealand participants (56.5%) taught at the primary level, followed by six participants at the secondary level. Half of the U.S. participants taught at the secondary level, followed by 13 participants at the elementary level. Primary/elementary teachers teach all academic subject areas in both the U.S. and New Zealand. Among the participants who taught at the intermediate/middle or secondary

school level, most of them taught English Language Arts/Literacy followed by Mathematics.

**Scenarios.** Table 4.4 below presents the mean and standard deviation for each of the 15 scenarios and the practice item. Scenario F456H (i.e., Facets 456 – High), which captures the high-level enactment, had the lowest mean 2.64 ( $SD = 0.63$ ), and Scenario F236L (i.e., Facets 236 – Low), which is a low-level scenario, had the highest mean 4.66 ( $SD = 0.69$ ). Based on the response categories and their corresponding values (i.e., *much lower* = 1, *slightly lower* = 2, *about the same* = 3, *slightly higher* = 4, *much higher* = 5), this suggests that the high-level scenarios were harder for participants to choose *about the same* or higher ratings and the low-level scenarios were easier for participants to choose *about the same* or higher ratings. The descriptive statistics of all scenarios seemed reasonable.

Table 4.4

*Descriptive Statistics for the Scenarios: Pilot Study*

<b>Item Entry Number/ Facets &amp; Level /Scenario Number</b>	<b>M</b>	<b>SD</b>
Entry #1 / Moderate level / Practice Item	3.55	0.88
Entry #2 / Facets 126 – Moderate / Scenario #2	3.64	1.01
Entry #3 / Facets 346 – High / Scenario #7	2.90	0.50
Entry #4 / Facets 156 – Low / Scenario #15	4.41	1.00
Entry #5 / Facets 236 – High / Scenario #4	2.88	0.50
Entry #6 / Facets 456 – Low / Scenario #12	3.90	0.81
Entry #7 / Facets 126 – Low / Scenario #3	4.44	0.73
Entry #8 / Facets 346 – Moderate / Scenario #8	3.62	0.76
Entry #9 / Facets 156 – High / Scenario #13	2.72	0.61
Entry #10 / Facets 236 – Low / Scenario #6	4.66	0.69
Entry #11 / Facets 456 – Moderate / Scenario #11	3.81	0.87
Entry #12 / Facets 126 – High / Scenario #1	2.67	0.60
Entry #13 / Facets 236 – Moderate / Scenario #5	3.63	0.79
Entry #14 / Facets 456 –High / Scenario #10	2.64	0.63
Entry #15 / Facets 346 – Low / Scenario #9	4.48	0.77
Entry #16 / Facets 156 – Moderate / Scenario #14	3.78	0.71

### **Rasch analyses**

The Rasch rating scale model (Andrich, 1978; Rasch, 1960/1980; Wright & Masters, 1982) was used to analyze the pilot data of 73 participants' responses to the 16 scenarios. I expected that the five scenarios capturing each of the three enactment levels (i.e., low, moderate, and high) would loosely form three clusters along the hierarchical continuum of the construct. Based on the empirical data (referring back to Table 4.4 above), the five low-level scenarios (i.e. Scenarios 3, 6, 9, 12, and 15) would be roughly located at the lower end of the continuum, the five moderate-level scenarios (i.e., Scenarios 2, 5, 8, 11, and 14) and the practice item would be in the middle, and the five high-level scenarios (i.e., Scenarios 1, 4, 7, 10, and 13) would be located at the upper end of the continuum. At this initial stage of scale development, the specific question was whether the Rasch analyses results demonstrated the *proof of concept*. In other words, would the scenario-style items capture the progression of enactment of practice for equity from the low, to the moderate, and to the high level of enactment?

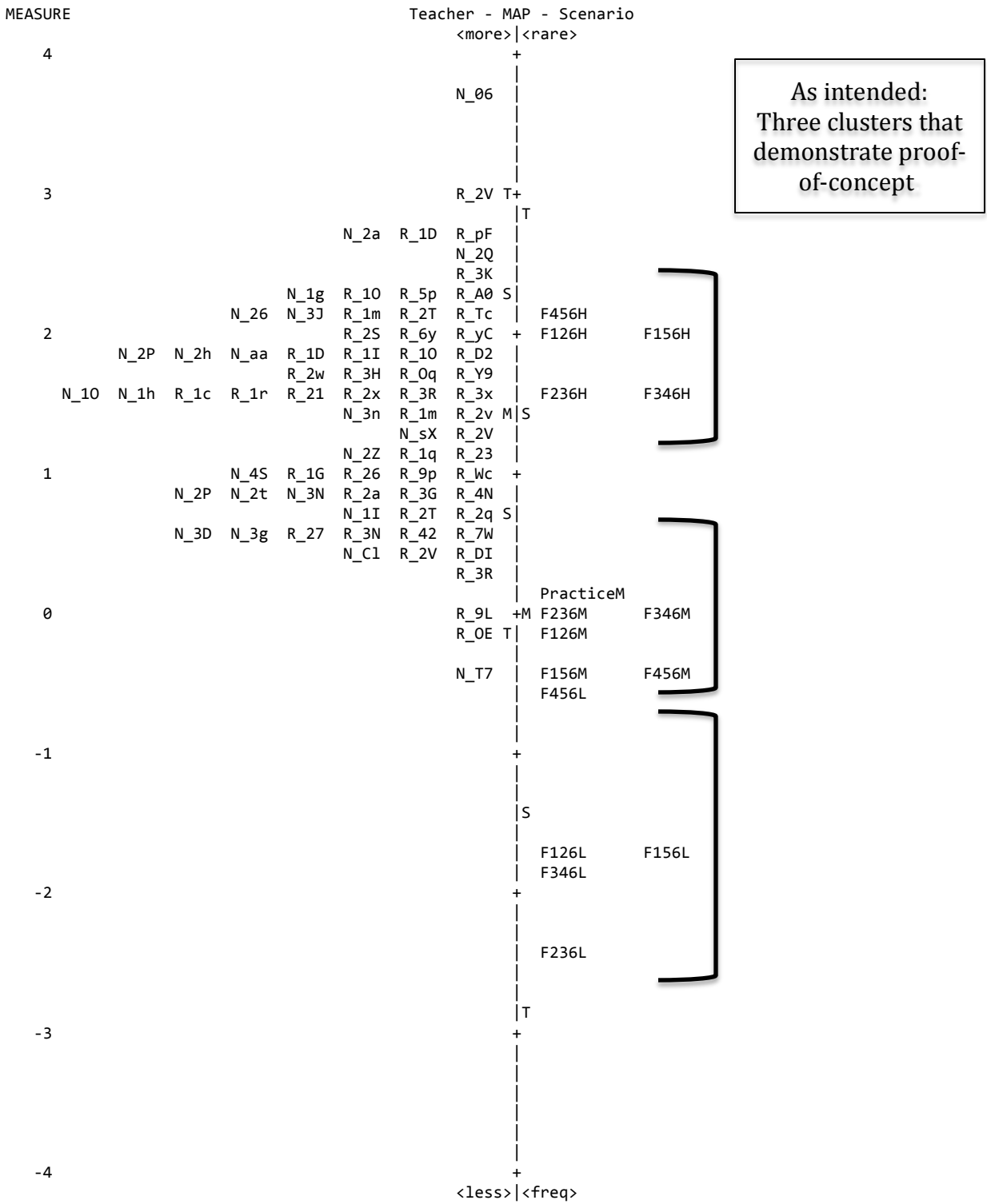
**Variable maps.** Figure 4.1 contains the variable map for the pilot study data, including several key elements. The central line of the variable map is a graphic representation of the scenarios and participating teachers' estimates along the hierarchical continuum of the construct – teachers' self-reports about their enactment of practice for equity. On the left-hand side of Figure 4.1 is the interval-level measurement in units of *logits* ranging from -4 to 4. The left-hand side of the central line indicates the locations of the participants (e.g., N\_06; N indicates New Zealand participants and R is for the U.S. participants) – their “enactment ability” estimates in *logits*. The right-hand side of the central line indicates the locations of scenario-style items (e.g., F456H; F456 indicates

the combination of Facets 4, 5, and 6 and the last letter indicates the levels of enactment - H = high, M = moderate, L = low) – their “enactment difficulty” estimates in *logits*. The difference between a respondent’s and an item’s location produces the probability of a person selecting a specific response to an item. The section of Rasch measurement models and key features in Chapter Two provided detailed explanations of *logits*. And, at the final stage of analysis, this form of the variable map has the logits re-expressed back into a raw score for the interpretation purposes.



Figure 4.1

Variable Map: Pilot Study



Overall, the results suggest that the locations of the scenarios followed the hypothesized order as shown in the variable map (See Figure 4.1). Starting at the bottom of the enactment scale, the five scenarios (i.e., Scenarios F126L, F236L, F346L, F456L, and F156L) were the easiest for participants to give “*higher*” ratings than the practice captured in the scenarios. In the middle of the distribution of scenarios were the five moderate-level items (i.e., Scenarios F126M, F236M, F346M, F456M, and F156M) and the moderate-level practice item. They are harder than the five low-level scenarios for the participants to give “*higher*” ratings to. These are followed by the five high-level scenarios (i.e., Scenarios F126H, F236H, F346H, F456H, and F156H), which were the hardest for participants to give “*higher*” ratings to than the practice captured in the scenarios. The order of the scenarios was consistent with my Rasch scale development expectation. In other words, moving up the scale means enacting from the lower to the higher and more complex practice for equity in the classroom as self-reported by the participants.

In addition to the item order, the locations of the items (i.e., the average difficulty estimates) tended to be lower than the locations of the participants (i.e., their “ability” estimates). The mean of participants’ levels of enactment (indicated by the “M” to the left of the variable map) was approximately 1.4 in *logits* and was higher than the mean of items set as zero in *logits* (the “M” to the right of the variable map). This suggests that the items tended to be easy for most participants to choose *about the same* or higher given the participants’ self-reported enactment levels.

While the distribution of the scenarios demonstrated the *proof of concept*, the scenarios were not equally spread, and there were gaps in the distribution of their

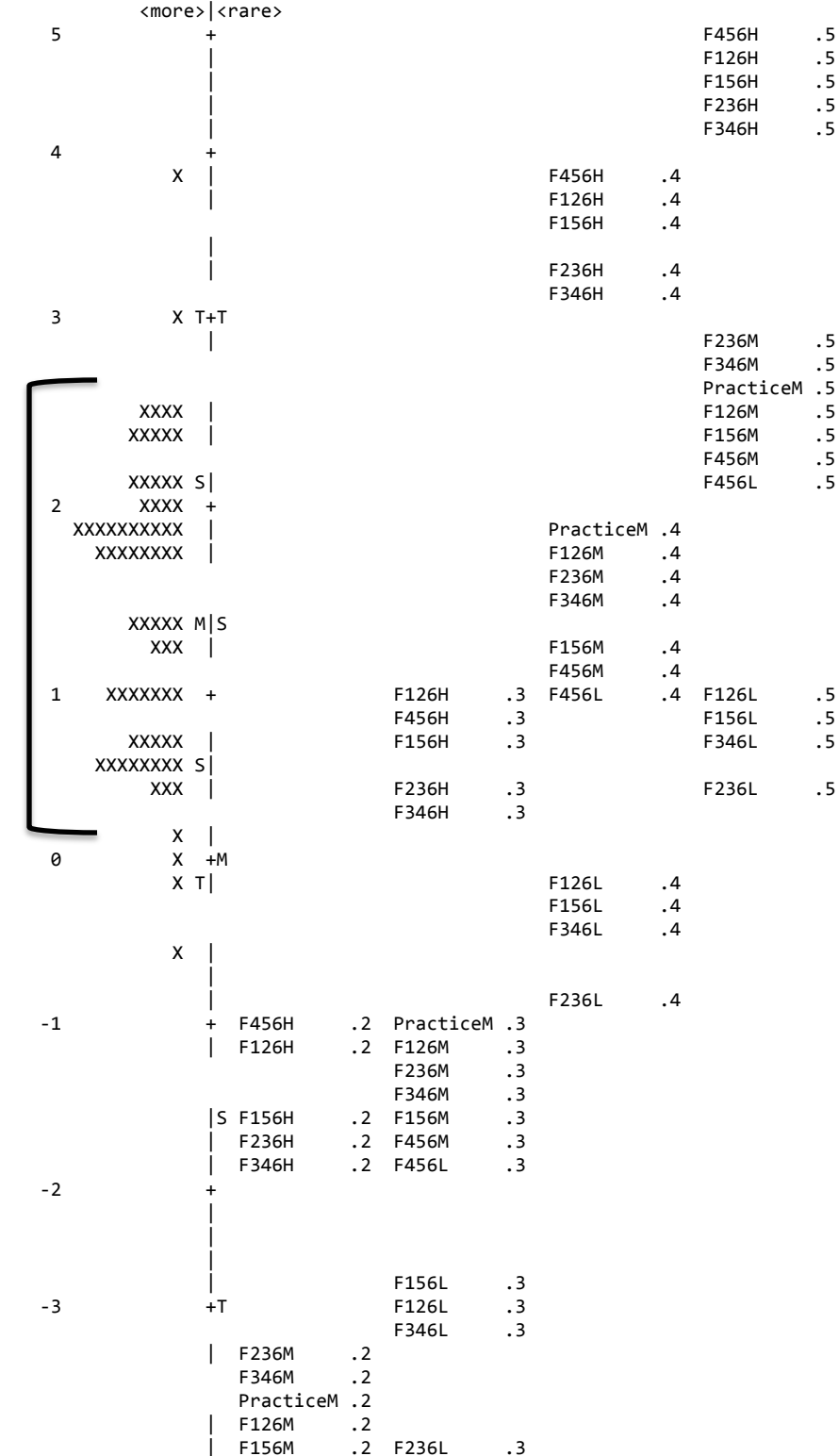
locations. Specifically, a large space exists between the two high-level scenarios (e.g., F236H, F346H) and the two moderate-level scenarios (e.g., F236M, F346M). There is also a space between the two moderate-level scenarios (e.g., F156M, F456M) and the two low-level scenarios (e.g., F126L, F156L). Therefore, to improve the scale, the next steps involve making some items even harder for participants to give positive ratings (i.e., *about the same* or higher) and addressing the gaps to make scenarios more equally distributed. In the Item revision section, I discuss the kinds of revisions made to improve the scale and the rationale.

Figure 4.2, below, presents similar information as Figure 4.1, but includes categorical threshold estimates. Figure 4.2 reveals the level of response expected of a respondent to any scenario item given their location. In this case, Andrich thresholds were used, which indicate the 50% probability of choosing one category or the next at the point (Andrich, 1978). As indicated in Figure 4.2, most participants were expected to respond *about the same* to the high-level scenarios, to choose *slightly higher* to the moderate-level scenarios and some low-level scenarios, and to select *much higher* to the low-level scenarios. No participants were expected to score *much higher* for the high-level scenarios.

Figure 4.2

*Cumulative Probability Map by Andrich Thresholds: Pilot Study*

MEASURE Teacher - MAP - Scenario - Andrich thresholds (modal categories if ordered)



*Cumulative Probability Map by Andrich Thresholds: Pilot Study (continued)*

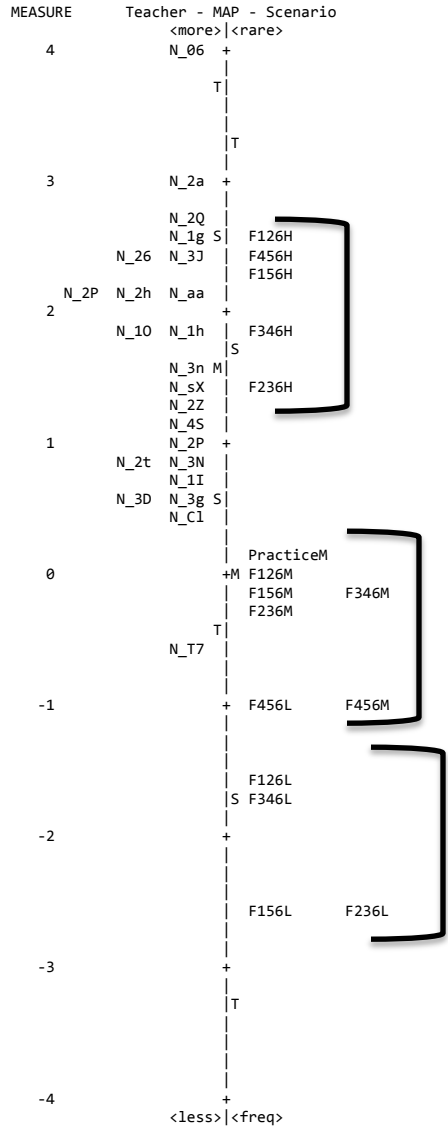
		F456M	.2
		F456L	.2
-4	+		
-5	+	F126L	.2
		F156L	.2
		F346L	.2
		F236L	.2
-6	+		
	<less> <freq>		

*Measurement invariance.* Since participants were recruited from two different cultural contexts (i.e., New Zealand and the United States) and had varying years of teaching experience (i.e., 44 participants taught for more than five years and 29 had less than five years of teaching experiences), I also examined whether the locations of the scenarios were identical, or nearly so, across subgroups. Figure 4.3 presents the variable maps for New Zealand and the U.S. participants separately, and Figure 4.4 presents the variable maps for novice (i.e., less than five years of teaching experience) and experienced teachers (i.e., five or more than five years of teaching experience) regardless of their country of origin. As shown in both figures, the order of the scenarios between the subgroups (i.e., New Zealand versus the United States participants and novice versus experienced participants) was fairly similar to each other and there was no serious disorder. Moreover, Pearson correlations between each of the two pairs of item difficulty estimates (see Table 4.5) were .965 (between New Zealand and the U.S. participants) and .979 (between the novice and experienced participants). This suggests that the meaning of enacting from lower to higher levels of teaching practice for equity in the classroom was invariant across the subgroups in terms of cultural contexts and years of teaching experience.

Figure 4.3

Variable Maps: Comparison between New Zealand and the U.S. participants

NZ Participants (N=23)



US Participants (N=50)

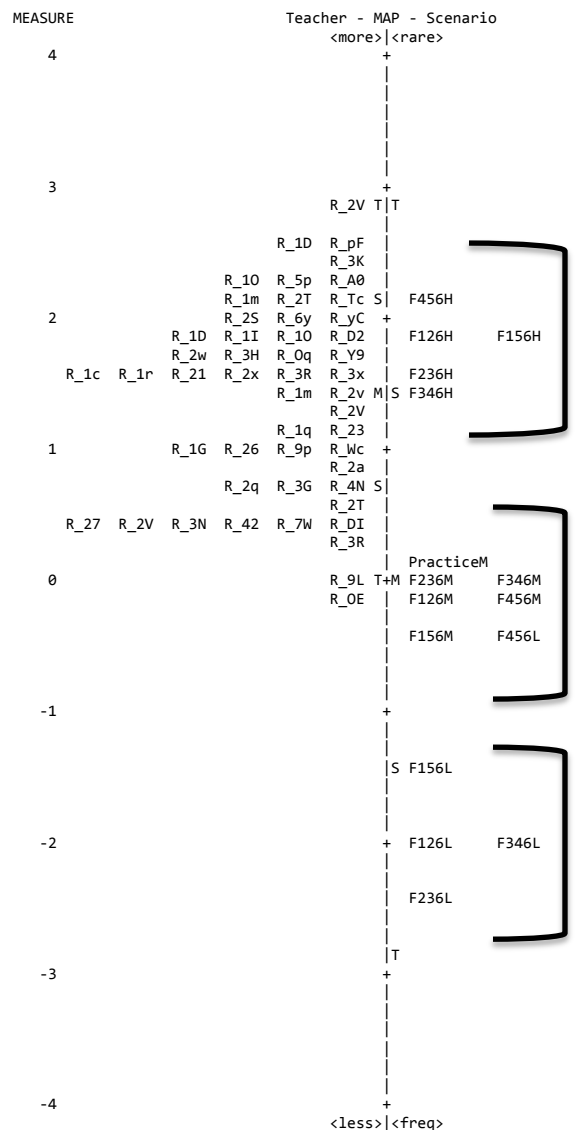
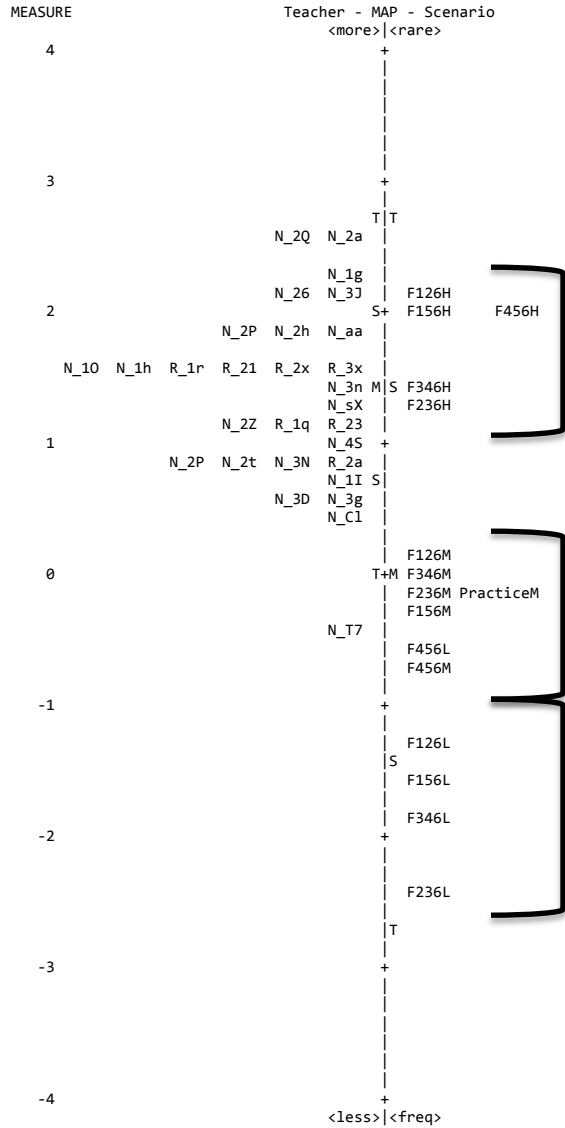


Figure 4.4

*Variable Maps: Comparison between novice and experienced participants*

**Novice Teachers (N=44)**



**Experienced Teachers (N=44)**

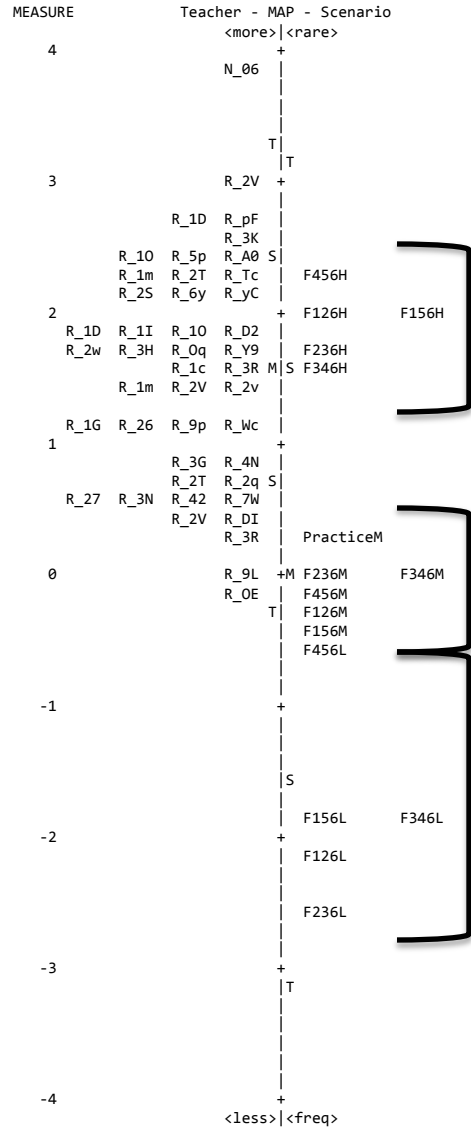




Table 4.5

*TEES Scenario Difficulty Estimates by Subgroups*

Scenarios	New Zealand (n=23)		U.S. (n=50)		Novice (n=29)		Experienced (n=44)	
	Difficulty	S.E.	Difficulty	S.E.	Difficulty	S.E.	Difficulty	S.E.
F126H	2.57	0.35	1.91	0.21	2.16	0.29	2.05	0.23
F236H	1.47	0.35	1.64	0.21	1.31	0.29	1.78	0.23
F346H	1.84	0.35	1.42	0.21	1.48	0.29	1.57	0.23
F456H	2.45	0.35	2.09	0.22	1.99	0.29	2.32	0.24
F156H	2.32	0.36	1.82	0.22	1.98	0.30	1.94	0.24
F126M	0.02	0.32	-0.13	0.20	0.13	0.27	-0.23	0.22
F236M	-0.27	0.31	0.03	0.20	-0.15	0.26	0.00	0.21
F346M	-0.17	0.31	0.03	0.20	-0.01	0.26	-0.05	0.21
F456M	-0.94	0.31	-0.18	0.20	-0.76	0.26	-0.15	0.22
F156M	-0.17	0.31	-0.43	0.20	-0.22	0.26	-0.46	0.21
F126L	-1.53	0.32	-1.94	0.24	-1.32	0.27	-2.19	0.27
F236L	-2.50	0.39	-2.46	0.28	-2.39	0.34	-2.52	0.30
F346L	-1.74	0.33	-1.99	0.24	-1.88	0.30	-1.92	0.26
F456L	-0.97	0.32	-0.47	0.20	-0.56	0.26	-0.64	0.22
F156L	-2.50	0.39	-1.42	0.22	-1.62	0.28	-1.79	0.25
	Pearson Correlation: 0.965				Pearson Correlation: 0.979			

**Goodness-of-fit analyses.** As mentioned in the previous two chapters, goodness-of-fit analyses provide empirical evidence on how well the data fit the Rasch model by identifying outliers and inconsistent response patterns for both items and persons. Table 4.6 below presents information on item misfit including item entry order, item difficulty measures in *logits*, weighted mean square of standardized residuals (Infit-MNSQ in WINSTEPS software) and the corresponding t statistics (Infit-ZSTD in WINSTEPS), and unweighted mean square of standardized residuals (Outfit-MNSQ in WINSTEPS) and the corresponding t statistics (Outfit-ZSTD in WINSTEPS). The weighted fit statistics are used to identify inconsistent response patterns, while the unweighted fit statistics are used to identify outliers. The reasonable range for the unweighted and weighted mean squares should be between 0.5 and 1.5 and a more liberal threshold of 1.2 is used to avoid missing potential problems in the initial stage (Ludlow et al., 2008; Ludlow et al., 2014). The absolute value of t statistics should range between  $\pm 2$  for a small sample size.

Most items had Infit statistics within the reasonable range between 0.5 and 1.5, as outlined in Table 4.6, below. Based on the liberal threshold of 1.2, four items (i.e., Scenarios F156L, F126M, F236L, and F346L) had Infit statistics greater than 1.2 (Infit-MNSQ = 1.93, 1.30, 1.29, and 1.26, respectively) and one item (i.e., F156L) had a t statistic (Infit-ZSTD)  $> 2$ . Among the four identified items, three (i.e., F156L, F126M, F346L) had Outfit statistics greater than 1.2 (Outfit-MNSQ = 1.74, 1.32, and 1.29, respectively) and Scenario F156L was the only item with a t statistic  $> 2$ . Given that most misfit items had negative logits and thus were “easier” items, the unexpected responses may come from high-scoring participants giving lower than expected responses to these items. The overall results of fit analysis suggest that the data fit the Rasch rating scale model well and that this scale measures teachers’ self-reports about their

enactment of practice for equity as a unidimensional construct with variation from low to high levels of enactment.

Table 4.6

*Item Misfit: Pilot Study*

Scenario	Entry Number	Item Difficulty	Infit		Outfit	
			MNSQ	ZSTD	MNSQ	ZSTD
F156L	4	-1.70	1.93	4.7	1.74	3.4
F126M	2	-0.09	1.30	1.9	1.32	1.9
F236L	10	-2.45	1.29	1.4	0.97	0.0
F346L	15	-1.89	1.26	1.5	1.29	1.4
PracticeM	1	0.10	1.20	1.3	1.17	1.0
F456M	11	-0.40	1.01	0.1	1.02	0.2
F456L	6	-0.61	0.99	0.0	0.99	0.0
F126L	7	-1.78	0.94	-0.3	0.90	-0.5
F156H	9	1.93	0.89	-0.6	0.88	-0.7
F456H	14	2.16	0.87	-0.7	0.86	-0.8
F236M	13	-0.06	0.79	-1.5	0.84	-1.1
F126H	12	2.07	0.83	-1.0	0.82	-1.0
F236H	5	1.58	0.71	-1.7	0.70	-1.8
F346M	8	-0.03	0.70	-2.2	0.71	-2.0
F346H	3	1.51	0.67	-2.0	0.67	-2.0
F156M	16	-0.36	0.64	-2.8	0.63	-2.8

To further investigate issues with each of the four misfit items and identify plausible explanations for them, I examined WINSTEPS Table 7.1 (the “Person Response Table” – not shown), which provides information about the misfit participants’ observed and expected responses, and standardized residuals. I also considered factors such as item entry order, items with missing value, and items’ levels of enactment measured.

The most poorly fitting item was Scenario F156L with both Infit and Outfit statistics greater than 1.2 and a t statistic greater than 2. Several poorly fitting participants gave unexpectedly low responses to this item with the magnitude of standardized residuals as large as 5. Similarly, Scenario F126M had both Infit and Outfit statistics greater than 1.2 (though the t

statistic falls within the range of  $\pm 2$ ). Several participants gave unexpectedly low responses to this item with standardized residuals of -2 and -3. While there was a practice item, which appeared prior to this scenario, Scenario F156M was technically the first item on the survey; hence, there may be a start-up effect on participants' responses.

Scenario F236L had an Infit statistic greater than 1.2, with three poorly fitting participants who gave unexpected low responses with standardized residuals ranging between -2 and -4. In terms of Scenario F346L that had both Infit and Outfit statistics greater than 1.2, only one poorly fitting participant gave a lower than expected rating to the scenario with a standardized residual of -5. This scenario was the only one to which this person had an unexpected response. This suggests that the misfit might be either random error or careless choice by this participant at the end of the scenario section. Therefore, I decided to exclude Scenario F346L from any further investigation and potential item revision.

As previously mentioned, six out of 73 participants skipped an item; therefore, I also examined whether the missingness occurred to any of the misfit items. However, there was no missing value for all four misfit items except for Scenario F126M which had one missing case. Last but not least, since all but one of the misfit items measured the low-level enactment of practice for equity, it is reasonable to suspect that some common issues may apply to these low-level scenarios.

To understand the specific issues residing in the three misfit scenarios (i.e., F156L, F126M, and F236L) and to obtain ideas about how to address them, I conducted a Think-aloud exercise with three key informants. In the following section, I briefly discuss the design of the Think-aloud exercise and detail the feedback received from them, the kind of revisions I made for the final full-scale administration, and the rationale for the revision.

### **Item revision**

As discussed in Chapter Three, the Think-aloud exercise involved having the key informants respond to a select number of good and misfit scenarios as well as a discussion session to detect the specific issues in the misfit items and ways to address them. In the case of this study, the key informants were asked to respond to five scenarios, including two good items (i.e., Scenarios F346H and F156M) and three misfit items (i.e., Scenarios F156L, F126M, and F236L). The informants were then encouraged to articulate their thought processes when responding to the three misfit items. Unlike the “poorly fitting” participants in the pilot study, all three informants gave responses that were as expected given their responses to the two good items. Therefore, they were asked to help identify the words, phrases, or sentences that could have potentially led the misfit participants to give unexpectedly low responses. Below, I discuss the issues identified for each of the three misfit items and the specific changes made to address them. I also discuss the revisions made to three items that I chose to improve the scale. Table 4.7 below presents the revisions made for these scenarios, and Table 4.8 shows the change of facet scores and the expected order for the final scale.

With regard to the first misfit item, Scenario F156L, all three informants pointed to the third sentence, *she follows standards and curriculum documents to design her lessons and make sure that students master and retain the content*, as the potential issue causing the misfit. Specifically, the verb *master and retain* could convey a higher level positive image than intended for the participants. To make the low-level description clearer, I changed the verb *follows* to *adheres to* and replaced *master and retain* with *memorize*. This change made the low-level description in Scenario F156L stronger but still kept a facet score of “3”.

The second misfit item was Scenario F126M. Two of the informants suggested that the last sentence of this moderate level scenario, *his explanations, however, are not always compelling and clear to all students and sometimes contain errors*, seemed to be less positive than and out of sync with the rest of the descriptions in the scenario. To make this scenario clearer and address a gap in the original scale (i.e., the gap between Scenarios F236H & F346H and Scenarios F236M & F346M), I changed the original sentence to *his explanations are clear and interesting to all students*. This change made Scenario F126M harder (an increase of a facet score from “6” to “7”) and I expected that this change would bring this scenario up to address the gap in the original scale. In addition, given that the “start-up” effect might be a reason for the misfit of this scenario, I adopted one informant’s suggestion to add a message between the practice item and this first scenario on Qualtrics. Specifically, the message would remind participants of the response option they had just chosen for the practice item after comparing their teaching against the description in the practice item. The purpose was to familiarize participants with the survey instruction and the cognitive process required to answer the scenarios.

The third misfit item was Scenario F236L. The informants indicated that the sentence, *Christine makes most of the decisions in the classroom and sets classroom expectation*, did not clearly capture a low-level of enactment. Therefore, I changed this sentence to *Christine, rather than involving the students, makes most of the decisions in the classroom*. I also revised the part *but generally* in the first sentence to *as* to make the low-level description clearer (but still a facet score of 3).

To stretch the upper boundary of the original scale, I revised the hardest scenario, Scenario F456H, to make it even more difficult to select *about the same* or higher ratings (but

still a facet score of 9). I rephrased the first sentence, *Juan has a strong sense of professional identity as a teacher and a strong commitment to advocating on behalf of students*, to *Juan is deeply committed to supporting the learning and life of diverse students, advocating on behalf of them, and contributing to the profession*. The revised sentence would convey a stronger stance and teaching practice with more specific description. With the same purpose in mind, I rephrased part of the third sentence (see Table 4.7) by adding positive adjectives (e.g., *constructive* and *timely feedback*). I also revised the last sentence by using the phrases such as *take charge of* and *experiment with* to convey an even more positive feeling.

To address the gap between Scenarios F236H & F346H and Scenarios F236M & F346M, I revised Scenario F346H to make this item easier, which lowered the facet score from “9” to “8”. To do so, I eliminated some of the positive-sounding adjectives throughout the scenario. The eliminated words included *genuinely*, *effectively*, and *consistently*. Lastly, I revised Scenario F456L in an attempt to address the gap between Scenarios F156M & F456M and Scenarios F126L & F156L (but still retain a facet score of 3). A second rationale for the revision was that Scenario F456L was located very close to Scenario F456M, but ideally should be lower. The revisions included using *executing* instead of *correctly implementing*; *attain* instead of *achieve*; and *solely* instead of *mainly*. Again, the rationale for these changes was that the new words or phrases delivered a clearer and more vivid image of the low-level enactment.

Table 4.7

*Item Revisions for the Final Study*

Scenarios	Revisions
<p><b>Scenario F156L (#15)</b> <b>Entry #4</b></p>	<p><u>Pilot version:</u> Adrian considers her role as a teacher primarily as transmitting knowledge to students. Adrian sets attainable goals for students but struggles to engage them. She <i>follows</i> standards and curriculum documents to design her lessons and makes sure that students <i>master and retain</i> the content. She often uses the same teaching strategies although she is unsure whether other approaches would be more or less effective for student learning. Adrian tends to work alone and sticks with what she knows.</p> <p><u>Final version:</u> Adrian considers her role as a teacher primarily as transmitting knowledge to students. Adrian sets attainable goals for students but struggles to engage them. She <i>adheres to</i> standards and curriculum documents to design her lessons and makes sure that students <i>memorize</i> the content. She often uses the same teaching strategies although she is unsure whether other approaches would be more or less effective for student learning. Adrian tends to work alone and sticks with what she knows.</p>
<p><b>Scenario F126M (#2)</b> <b>Entry #2</b></p>	<p><u>Pilot version:</u> Tim holds high expectations for some students in his class and mostly communicates these expectations clearly. He generally sees students' home culture as a strength and collaborates with some parents/caregivers. He sometimes lets his students choose a topic consistent with their interests to further their learning. Tim sometimes draws on cultural examples to design learning experiences that are relevant to students. He utilizes a selected number of approaches to explain key concepts. His explanations, <i>however, are not always compelling and clear to all students and sometimes contain errors.</i></p> <p><u>Final version:</u> Tim holds high expectations for some students in his class and mostly communicates these expectations clearly. He generally sees students' home culture as a strength and collaborates with some parents/caregivers. He sometimes lets his students choose a topic consistent with their interests to further their learning. Tim sometimes draws on cultural examples to design learning experiences that are relevant to students. He utilizes a selected number of approaches to explain key concepts. His explanations <i>are clear and interesting to all students.</i></p>
<p><b>Scenario F236L (#6)</b></p>	<p><u>Pilot version:</u> Christine occasionally engages with parents but generally she sees this as unnecessary. She has a quiet, reserved manner with her students, approaching all students the same way. Christine <i>makes most of the</i></p>



<p><b>Entry #10</b></p>	<p><i>decisions in the classroom and sets classroom expectations.</i> Because she believes that students should work individually, Christine does not see the need to facilitate interactive skills and usually assigns work for students to carry out on their own. Her classroom is arranged to be a generally quiet, non-interactive space.</p> <p><u>Final version:</u> Christine occasionally engages with parents but generally she sees this as unnecessary. She has a quiet, reserved manner with her students, approaching all students the same way. Christine, <i>rather than involving the students, makes most of the decisions in the classroom.</i> Because she believes that students should work individually, Christine does not see the need to facilitate interactive skills and usually assigns work for students to carry out on their own. Her classroom is arranged to be a generally quiet, non-interactive space.</p>
<p><b>Scenario F456H (#10)</b></p> <p><b>Entry #14</b></p>	<p><u>Pilot version:</u> <i>Juan has a strong sense of professional identity as a teacher and a strong commitment to advocating on behalf of students.</i> Juan builds on students' perspectives and draws on a variety of sources to identify learning priorities and teaching strategies. He involves students in designing assessments and fully integrates assessment into his instruction to <i>evaluate his own practice and to give feedback to students.</i> Juan <i>continuously reflects on his practice and tries out new approaches to motivate and respond to students' learning needs.</i></p> <p><u>Final version:</u> Juan is <i>deeply committed to supporting the learning and life of diverse students, advocating on behalf of them, and contributing to the profession.</i> Juan builds on students' perspectives and draws on a variety of sources to identify learning priorities and teaching strategies. He involves students in designing assessments and fully integrates assessment into his instruction to <i>provide constructive and timely feedback to students.</i> Juan <i>takes charge of his professional learning through continuous reflection on his practice and experimenting with new approaches to motivate and respond to students' learning needs.</i></p>
<p><b>Scenario F456L (#12)</b></p> <p><b>Entry #6</b></p>	<p><u>Pilot version:</u> Dave believes good teaching is <i>correctly implementing</i> a set of techniques to ensure that students <i>achieve</i> curriculum expectations for their grade/year levels. He <i>mainly</i> relies on standards or curriculum documents to identify learning priorities and teaching approaches. He designs assessments on his own and generally uses them to check whether students meet the minimum academic standards. Dave reviews his students' test results, sometimes altering his practice to boost their scores.</p> <p><u>Final version:</u> Dave believes good teaching is <i>executing</i> a set of techniques to ensure that students <i>attain</i> curriculum expectations for their grade/year levels. He <i>solely</i> relies on standards or curriculum documents to identify</p>

	learning priorities and teaching approaches. He designs assessments on his own and generally uses them to check whether students meet the minimum academic standards. Dave reviews his students' test results, sometimes altering his practice to boost their scores.
<b>Scenario F346H (#7)</b> <b>Entry #3</b>	<p><u>Pilot version:</u> Katherine <i>genuinely</i> cares for and respects her students. She encourages students to be independent learners and to investigate and build understanding of their own, and she involves them in setting criteria and goals for their learning. She <i>effectively</i> constructs her teaching practice to be engaging to all students, and integrates a variety of assessment approaches well into her teaching. Katherine interacts with students to provide constructive feedback and adjusts her practice appropriately. She <i>consistently</i> monitors and facilitates collaborative learning among her students.</p> <p><u>Final version:</u> Katherine cares for and respects her students. She encourages students to be independent learners and to investigate and build understanding of their own, and she involves them in setting criteria and goals for their learning. She constructs her teaching practice to be engaging to all students, and integrates a variety of assessment approaches well into her teaching. Katherine interacts with students to provide constructive feedback and adjusts her practice appropriately. She monitors and facilitates collaborative learning among her students.</p>

I incorporated all revisions mentioned above in the final survey (see Appendix J on page 292) and administered the survey on Qualtrics again. Table 4.8 summarizes what changes were made for the final survey and what remained the same, as well as the expected order of the scenarios (see Appendix L on page 315 for revised scenarios in “difficulty” order). The results of the analyses for the final full-scale administration are presented in the following section.

Table 4.8

*Revised Sentence Mapping Levels: Changes from Pilot to Final Survey*

Level of Enactment	Scenario	Facet						Total Score	
		1	2	3	4	5	6	Pilot	Final
High (Level 3)	F126H (#1)	X	X				X	9	9
	F236H (#4)		X	X			X	9	9
	F346H (#7)			X	X		X	9	8 (Revised to be easier with a lower score)
	F456H (#10)				X	X	X	9	9 (Revised to be the hardest)
	F156H (#13)	X				X	X	9	9
Moderate (Level 2)	F126M (#2)	X	X				X	6	7 (Revised to be harder with a higher score)
	F236M (#5)		X	X			X	6	6
	F346M (#8)			X	X		X	6	6
	F456M (#11)				X	X	X	6	6
	F156M (#14)	X				X	X	6	6
Low (Level 1)	F126L (#3)	X	X				X	3	3
	F236L (#6)		X	X			X	3	3 (Revised for misfit)
	F346L (#9)			X	X		X	3	3
	F456L (#12)				X	X	X	3	3 (Revised to be easier)
	F156L (#15)	X				X	X	3	3 (Revised for misfit)

## **Final Full-Scale Administration**

### **Overview of participants' responses**

The pool of participants for the final full-scale administration included pre-service teachers (enrolled in a full practicum in the fall semester of 2016) and novice teachers (graduated within past 1-3 years from teacher preparation programs at Boston College). I also recruited participants (i.e., teacher candidates currently enrolled in full-practicum or student teaching, recent program graduates, or in-service teachers) through ten teacher educators at other higher education institutions.

Ninety-seven participants in total responded to the survey. Twenty-three pre-service teachers from Boston College responded out of 60 who were contacted via email and newsletters, which yielded a response rate of 38.3%. Thirty-eight recent program graduates from Boston College responded out of 147 who were recruited by multiple emails (a response rate of 25.8%). In addition, 36 participants recruited through the ten teacher educators at other universities responded the survey.

**Missing data.** The data in Table 4.9 represent the number of participants based on the number of scenarios they skipped. Among the 97 participants who responded to the survey, 60 (61.9%) completed the 16 scenarios, and all except one participant also completed the rest of the survey. Three participants (3.1%) skipped one scenario but completed the rest of the survey. All participants who skipped six or more scenario items (34 out of 97, 35.1%) left the survey completely at some point, and none of them responded to items in the following two sections of the survey. Although 97 respondents agreed to participate in the survey, 24 (24.7% out of 97) dropped out when they saw the survey instructions and the practice item on the following page of the survey. This missing data pattern – a steep drop in response rate once participants

encountered the unconventional nature of the scenario items - was very similar to the pilot study data.

Table 4.9

*Number Missing: Final Study*

<b>Number of Missing Variables</b>	<b>Frequency (Number of Cases)</b>	<b>Percent</b>	<b>Cumulative Percent</b>
0	60	61.9	61.9
1	3	3.1	64.9
6	1	1.0	66.0
9	3	3.1	69.1
12	1	1.0	70.1
15	5	5.2	75.3
16	24	24.7	100.0
<b>Total</b>	97	100.0	

I also examined the missing cases by sample sources and responses to the practice item as presented in Tables 4.10 and 4.11, respectively. In terms of sample sources, the dropout rate of both pre-service teachers and graduates at Boston College was approximately 25.0%, while the dropout rate among participants from other institutions was 55.5% (20 out of 36). A plausible explanation may be that my professional affiliation to those participants prepared by the programs at Boston College had a positive impact on the completion rate. With regard to the dropout participants' response choices to the practice item, most selected *about the same* and some chose *slightly higher*. This was very similar to the pattern observed in the pilot study. While *about the same* and *slightly higher* were the two most frequently used response options by all respondents, it was likely that the dropout participants were unsure how to respond given the multiple-sentence nature of the item. Based on the results of examining the missing data, I used the *Listwise Deletion Method* again and kept participants who had only one missing value. This decision led to a total of 63 participants (64.9% of 97 respondents) in the item analysis.

Table 4.10

*Number Missing by Sample Sources: Final Study*

<b>Number of Items Missing</b>	<b>BC students</b>	<b>BC graduates</b>	<b>Non-BC participants</b>	<b>Total</b>
0	17	27	16	60
1	1	1	1	3
6	0	0	1	1
9	0	0	3	3
12	0	0	1	1
15	0	1	4	5
16	5	9	10	24
<b>Total</b>	<b>23</b>	<b>38</b>	<b>36</b>	<b>97</b>

Table 4.11

*Number Missing by Response Categories of the Practice Item: Final Study*

<b>Number Missing</b>	<b>Response Category</b>					<b>Total</b>
	<b>Much lower</b>	<b>Slightly lower</b>	<b>About the same</b>	<b>Slightly higher</b>	<b>Much higher</b>	
<b>0</b>	1	3	25	25	6	60
<b>1</b>	1	0	1	1	0	3
<b>6</b>	0	0	0	1	0	1
<b>9</b>	0	1	1	1	0	3
<b>12</b>	0	0	1	0	0	1
<b>15</b>	0	0	4	1	0	5
<b>Total</b>	<b>2</b>	<b>4</b>	<b>32</b>	<b>29</b>	<b>6</b>	<b>73</b>

## **Descriptive statistics**

*About the participants.* Among the 63 participants in the analysis, the majority of participants were female (79.3%). There were ten male participants and two participants who identified themselves as *Other*. With regard to participants' race and ethnicity, 51 participants (80.9%) identified as White, five identified as Asian, followed by three participants with Hispanic, Latino, or Spanish origin, two Black or African American, and one with Other origins.

Similar to the composition of the pilot study participants, participants in the final full-scale administration also had varying years of teaching experience. 19 participants (30.1%) reported having less than one year of experience, 17 participants (26.7%) had between one to three years of teaching experience, and 17 participants (26.7%) had between three to five years of teaching experience. Four participants reported teaching for between five and ten years, and five participants taught for more than 10 years.

With regard to the grade-level and subject areas that participants primarily taught, the majority of participants (32 out of 63, 50.75%) taught at the elementary level, 21 participants (33.3%) taught at the secondary level, eight participants (12.7%) taught at the middle school level, and one at the college level. In terms of the subject areas, 17 participants (26.9%) taught English Language Arts followed by six participants teaching history/social studies, eight teaching mathematics, and eight teaching science/engineering. Among the 21 participants who reported *Other* to the question on subject area, five participants were special education teachers. Given the criteria of target participants, I decided to exclude the five special education teachers and one teacher who taught at the college level. Following these exclusions, a total of 57 participants remained for the Rasch analyses.

**Scenarios.** Table 4.12 below presents the mean and standard deviation for each of the 15 scenarios and the practice item. Scenario F456H had the lowest mean 2.65 ( $SD = 0.69$ ) and Scenario 236L the highest mean 4.53 ( $SD = 0.93$ ). Given the response categories and corresponding values (i.e., *much lower* = 1, *slightly lower* = 2, *about the same* = 3, *slightly higher* = 4, *much higher* = 5), this means that the high-level scenarios were harder for participants to select *about the same* or higher ratings and the low-level scenarios were easier for the participants to give *about the same* or higher ratings. The descriptive statistics for the scenarios were reasonable.

Table 4.12

*Descriptive Statistics for the Scenarios: Final Study*

<b>Item Entry Number/ Facets &amp; Level /Scenario Number</b>	<b>M</b>	<b>SD</b>
Entry #1 / Moderate level / Practice Item	3.49	0.89
Entry #2 / Facets 126 – Moderate / Scenario #2	3.23	0.68
Entry #3 / Facets 346 – High / Scenario #7	2.98	0.55
Entry #4 / Facets 156 – Low / Scenario #15	4.35	1.08
Entry #5 / Facets 236 – High / Scenario #4	2.93	0.49
Entry #6 / Facets 456 – Low / Scenario #12	4.04	0.93
Entry #7 / Facets 126 – Low / Scenario #3	4.39	0.94
Entry #8 / Facets 346 – Moderate / Scenario #8	3.63	0.75
Entry #9 / Facets 156 – High / Scenario #13	2.71	0.78
Entry #10 / Facets 236 – Low / Scenario #6	4.53	0.93
Entry #11 / Facets 456 – Moderate / Scenario #11	3.96	0.71
Entry #12 / Facets 126 – High / Scenario #1	2.73	0.67
Entry #13 / Facets 236 – Moderate / Scenario #5	3.61	0.73
Entry #14 / Facets 456 –High / Scenario #10	2.65	0.69
Entry #15 / Facets 346 – Low / Scenario #9	4.30	1.18
Entry #16 / Facets 156 – Moderate / Scenario #14	3.81	0.88

**Rasch analyses**

The Rasch rating scale model was used to test whether the revised scenarios followed the expected order as specified earlier in Table 4.8. Specifically, I expected that the revision would result in less clustering and a more equal uniform spread of scenarios compared to the earlier



distribution of the scenarios based on the pilot study data. I also compared the separation indices for both persons and items between the pilot and the final full-scale administration, and investigated the goodness-of-fit for both poorly fitting items and persons. Since the purpose of this study was to develop a scale that measures the enactment of practice for equity in a meaningful and authentic manner, I deleted five poorly fitting persons based on conservative fit analysis thresholds to obtain a more productive measurement as suggested in the Rasch literature (Linacre, 2016; Smith, Linacre, & Smith, 2003). I discuss the rationale for choosing the misfit respondents for deletion and present the final psychometric results on the data of 52 participants, including the final variable maps with descriptors for raw scores, categorical characteristic curves, goodness-of-fit analyses, effect of social desirability, and residual analysis.

*Variable maps.* To recap, Figure 4.5 contains the original distribution of the scenarios for the pilot study. Based on the empirical data of 73 participants in the pilot study, the locations of the scenarios along the hierarchical continuum of the enactment construct (i.e., presented as the central line in the figure) confirmed the *a priori* theory. That is, the five high-level scenarios were expected to be harder to give higher ratings to than the five moderate-level scenarios and the five moderate-level scenarios were expected to be harder to give higher ratings to than the five low-level scenarios, which were hypothesized to be the easiest to rate high. The scale structure of the three loose clusters of items defining low to moderate to the high level of enactment provided a *proof of concept*. However, as discussed earlier, the scenarios could be more equally spread and be more on target with the participants through stretching the upper boundary of the scale. To improve the scale, I employed the same measurement strategy described in the Procedure section of Chapter Three to revise four scenarios (i.e., Scenarios F456H, F346H, F126M, and F156L).

Specifically, I revised Scenario F456H to be the hardest scenario (retaining a facet score of “9”) with the intention to stretch the upper boundary of the original scale. To address gaps between the three clusters of scenarios, I also revised Scenario F346H to be easier resulting in a lowered score of “8” (from “9”), Scenario F126M to be harder resulting in a higher score of “7”(from “6”), and Scenario F156L to be easier (retaining its score of “3”).

Figure 4.6 presents the variable map for the final full-scale administration. As shown in Figure 4.6, the distribution of the revised scenarios reflected, in general, what was expected. First, while Scenario F456H was the hardest item among all the 15 scenarios, the revision did not seem to stretch the upper boundary of the original scale as much as expected when comparing Figures 4.5 and 4.6. Scenario F346H was expected to be easier to give higher ratings to than the rest of the high-level scenarios, and the location was lowered as intended. Scenario F126M was expected to be harder than the rest of the moderate-level scenarios, and it did move up along the scale as intended. Both revised scenarios filled the gap between Scenarios F236H & F346H and Scenarios F236M & F346M in the original distribution of the scenarios. Lastly, while the revised Scenario F456L moved down the scale and was located below Scenario F456M, the revision did not seem to widen the distance between the two scenarios as much as intended. The gap between Scenarios 156M & 456M and Scenarios F126L & F156L in the original scale from the pilot study, however, was not present in the final administration (See Figure 4.6).

Figure 4.5

Variable Map: Pilot Study

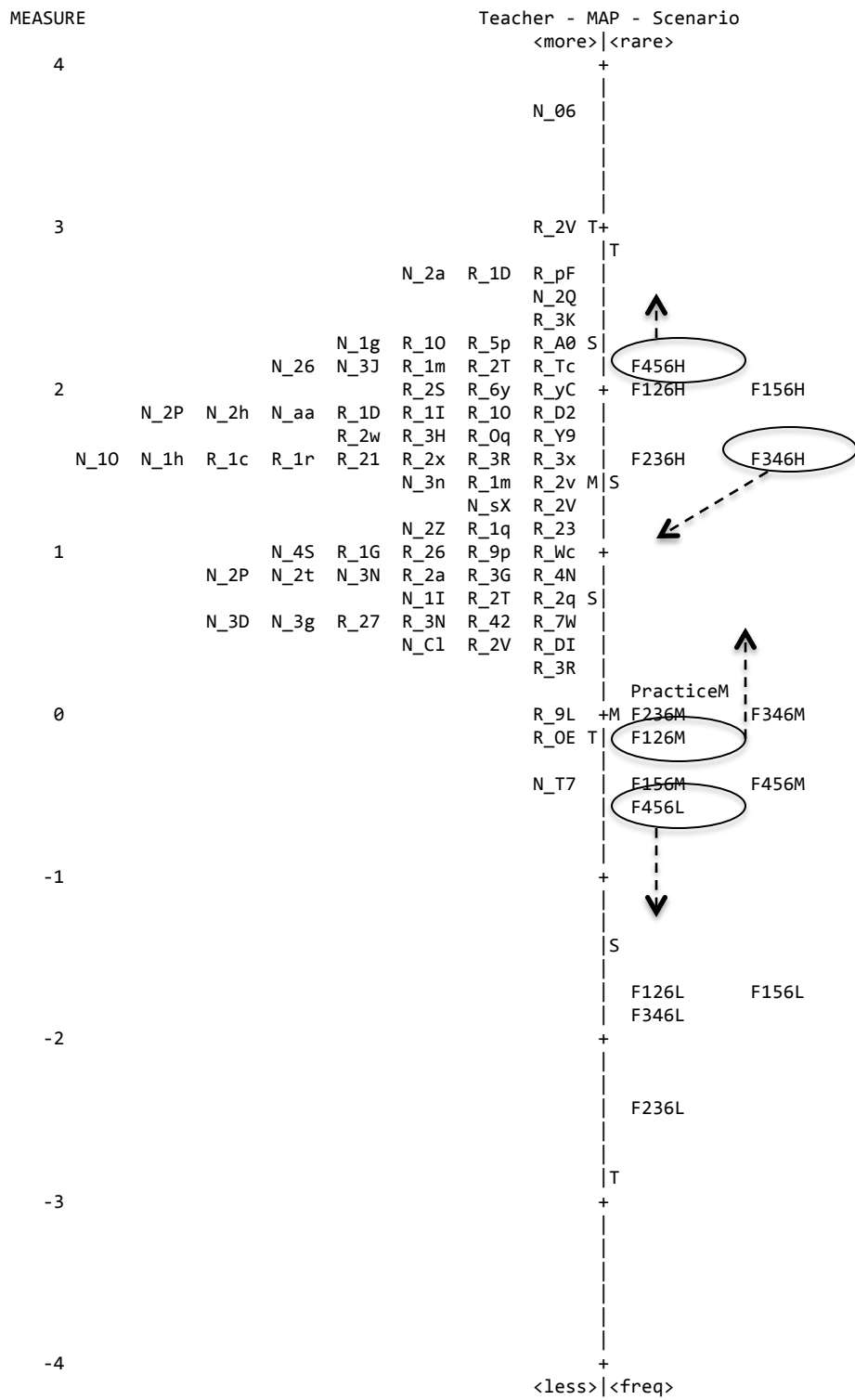
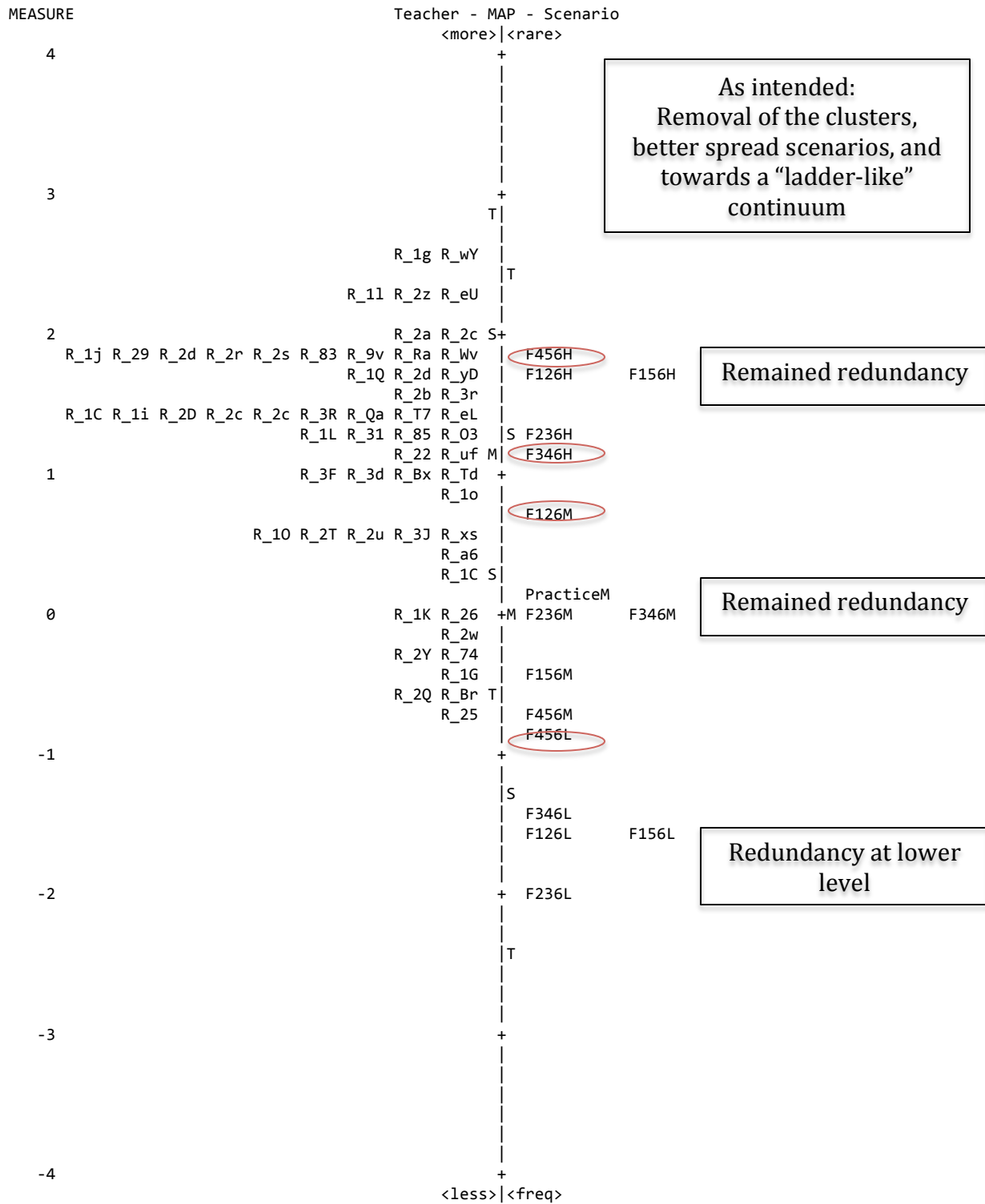


Figure 4.6

Variable Map: Final Study (57 Cases)

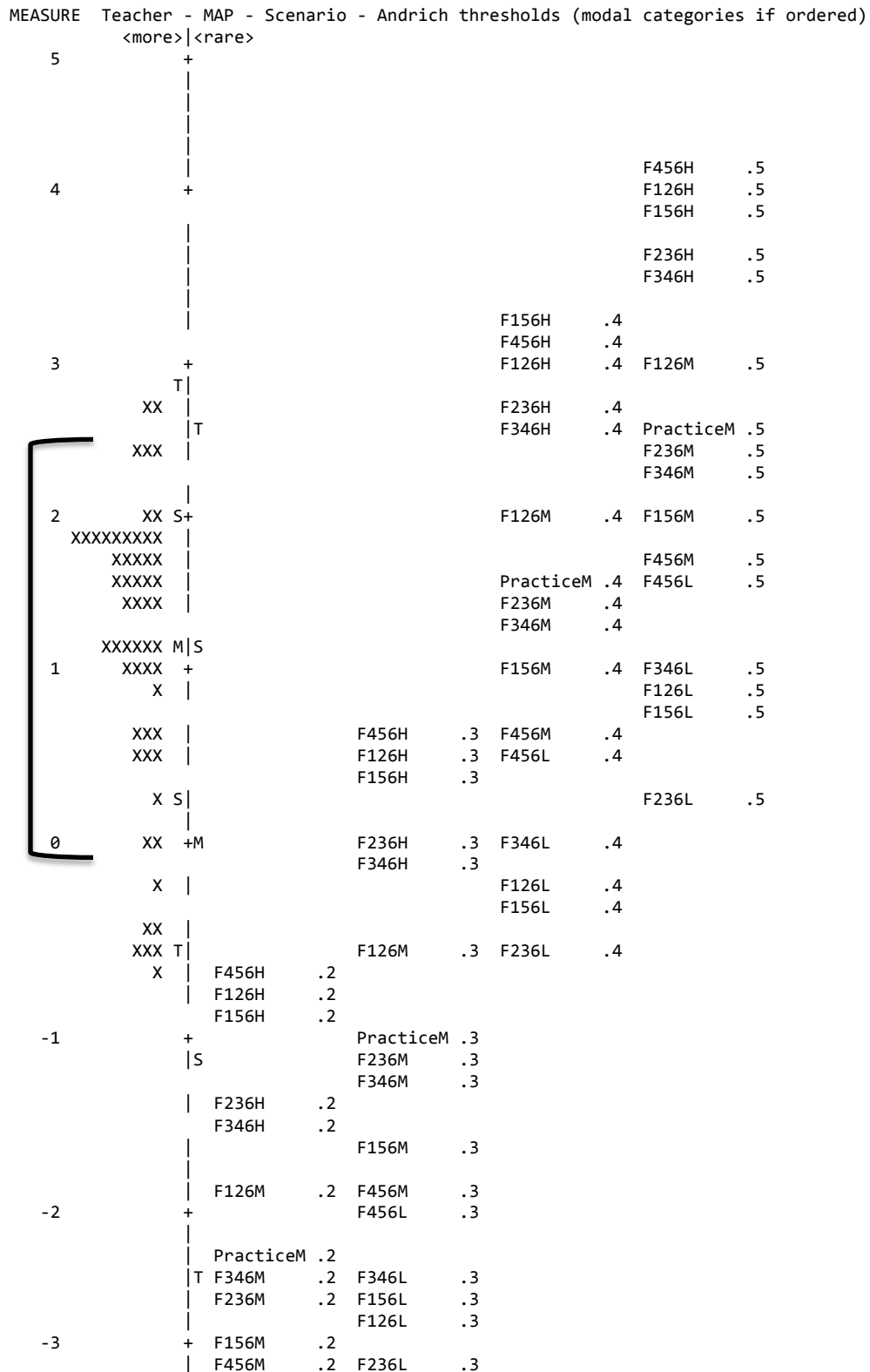


Overall, the revised scenarios became more sufficiently and reasonably spread to define the increasing levels of enactment of practice for equity compared to the distribution of the original pilot scenarios. The person separation was 1.76 (a slight increase from 1.63 in the pilot study) with a person reliability of .76. The item separation was 5.70 (a decrease from 7.51 in the pilot study) with an item reliability of .97. The item separation decreased from the pilot study, though it was still larger than 3.0. In addition, while the enactment level of the participants remained higher than the difficulty levels of the items, the locations of the revised scenarios seemed to be more on target with the participants compared to the original scale as indicated by the slight increase of person separation.

Figures 4.7 below represents variable maps that present information that is similar to Figure 4.6, but now item difficulties are separated by Andrich thresholds which indicate the 50% probability of moving from one category to the next at that point. As shown in Figure 4.7, the majority of participants were expected to select *about the same* to the high-level scenarios and to choose *slightly higher* to the moderate-level scenarios.

Figure 4.7

*Cumulative Probability Map by Andrich Thresholds: Final Study (57 Cases)*



Cumulative probability map by Andrich thresholds: Final full-scale (57 cases)

(continued)

		F456L	.2
-4	+	F156L	.2
		F346L	.2
		F126L	.2
		F236L	.2
-5	+		
	<less> <freq>		

**Goodness-of-fit.** Table 4.13 below contains information for item misfit, including both weighted and unweighted mean squares of standardized residuals between the expected and observed responses and the t statistics associated with it. As mentioned previously, the reasonable range for the unweighted (Outfit-MNSQ) and weighted (Infit-MNSQ) mean square should fall between 0.5 and 1.5 (Linacre, 2016). The t statistic range should fall between  $\pm 2$  for a small sample size.

A liberal threshold of 1.3 was used to identify items that were outliers (i.e., Outfit-MNSQ > 1.3) or produced inconsistent responses (i.e., Infit-MNSQ > 1.3). Four misfit items were identified, including Scenarios F156L, F346L, F126L, and F156H. These items had both the weighted mean square (Infit-MNSQ = 1.76, 1.79, 1.45, and 1.61, respectively) and unweighted mean square (Outfit-MNSQ = 1.82, 1.56, 1.61, and 1.60, respectively) greater than 1.3. All four items had t statistics (ZSTD) for both Infit- and Outfit-MNSQ larger than 2. Three misfit items - Scenarios F156L, F346L, and F126L - had low item difficulty measures (i.e., negative *logits*) and were easier items. And, Scenario F156L and Scenario F346L were misfit items in the pilot study and remained so after the revision. For low-level misfit items, the misfit may be the result of high-scoring participants who gave a lower than expected responses. For the high-level misfit item (i.e., Scenario F156H), the misfit may be the result of low-scoring participants who gave a higher than expected response.



Table 4.13

*Item Misfit: Final Study*

Scenario	Entry Number	Item Difficulty	Infit		Outfit	
			MNSQ	ZSTD	MNSQ	ZSTD
F156L	4	-1.50	1.76	3.4	1.82	3.3
F346L	15	-1.38	1.79	3.6	1.56	2.5
F126L	7	-1.59	1.45	2.1	1.61	2.5
F156H	9	1.71	1.61	2.7	1.60	2.6
F236L	10	-1.96	1.32	1.5	0.92	-0.2
F126H	12	1.65	1.25	1.2	1.25	1.2
F456L	6	-0.82	1.14	0.8	1.13	0.8
PracticeM	1	0.18	0.92	-0.4	0.94	-0.3
F456H	14	1.81	0.86	-0.7	0.85	-0.8
F126M	2	0.67	0.73	-1.6	0.71	-1.6
F236H	5	1.24	0.72	-1.5	0.69	-1.7
F156M	16	-0.39	0.65	-2.4	0.66	-2.2
F346H	3	1.14	0.60	-2.3	0.58	-2.4
F346M	8	-0.03	0.58	-2.9	0.60	-2.6
F456M	11	-0.69	0.51	-3.6	0.53	-3.4
F236M	13	-0.04	0.52	-3.4	0.52	-3.3

Once the misfit items were identified, I examined the person fit statistics together with person response output table in WINSTEP (Table 7.1 – not shown) to understand factors that may contribute to the unexpected responses. Table 4.14 below presents the weighted (i.e., Infit-MNSQ) and unweighted (i.e., Outfit-MNSQ) mean square of standardized residuals and the t statistics (ZSTD) associated with the mean squares for the participants. Table 4.14 includes only those participants with both the Infit-MNSQ and Outfit-MNSQ larger than 1.3. Further investigation into the unexpected responses focused on the first five participants with both mean squares larger than 1.6 and a ZSTD larger than 3.0.

Table 4.14

*Person Misfit: Final Study*

Participants	Person Ability	Infit		Outfit	
		MNSQ	ZSTD	MNSQ	ZSTD
#27	1.88	5.89	6.9	8.15	8.1
#51	-0.72	5.76	6.8	5.66	6.7
#35	-0.31	4.60	6.0	4.38	5.7
#57	-0.55	3.45	4.6	3.39	4.5
#45	0.28	2.32	3.0	2.25	2.9
#17	1.36	1.67	1.7	1.73	1.8
#16	0.04	1.58	1.6	1.57	1.5
#12	0.63	1.41	1.2	1.39	1.1
#3	-0.55	1.37	1.1	1.37	1.1

The first misfit participant (#27) was a high-scoring respondent (ability measure 1.88) who gave lower than expected responses to three of the low-level scenarios (i.e., Scenario F156L, F126L, and F456L) with standardized residuals as large as -6 and -7 for the first two items and -2 for Scenario F456L. It is possible that the two low-level scenarios (i.e., F156L and F126L, both misfit) captured an image of teaching so unfavorable that this participant immediately chose *Much lower* without thinking carefully of the survey instruction. Rather than comparing his/her own teaching practice against the practice in the scenarios, this participant responded the opposite way. The fact that this participant responded *Slightly higher* or *Much higher* to most items might offer further evidence to support this hypothesis.

Participants #51 and #35 were both low-scoring respondents (ability measures -0.72 and -0.31, respectively) and had similar response patterns. Participants #51 and #35 gave even lower than expected responses to the low-level scenarios (i.e., Scenario F156L, F346L, and F126L – all misfit) with standardized residuals of -3. Interestingly enough, participants #51 and #35 also tended to give higher than expected responses to the high-level scenarios such as Scenario

F156H (a misfit) with standardized residuals of 4 and 2 respectively. It is possible that the two participants were not clear with the survey instructions (i.e., compare your practice against the practice in the scenarios and choose the response). Therefore, they chose lower than expected responses to low-level scenarios because the teaching practice described in these scenarios was “lower” than what they do; and, they chose higher than expected responses to the high-level scenarios because the teaching practice in those scenarios was “higher” than their practice.

Participant #57 had a low score (ability measure -0.55) and tended to give lower than expected responses to low-level scenarios such as Scenarios F346L (a misfit), F236L, and F456L. Again, the unexpected responses could be due to the unfavorable images of teaching practice that the low-level scenarios capture, which may have prompted the participant to choose *Much lower* without comparing his/her practice against the practice in the scenarios. In addition, this participant gave the same response category *About the same* to a series of items towards the end of the scenario section (i.e., four items with entry order from 11 to 14). While these items were not misfit and the responses were not particularly unexpected for these four items, this response pattern might imply increasing levels of fatigue or carelessness for this participant, supported by the fact that this participant did not continue answering items in later sections of the survey.

Lastly, participant #45 had a person location estimate of 0.28. This participant gave lower than expected responses to two low-level scenarios (misfit scenario F156L and Scenario F236L) and higher than expected responses to two high-level scenarios (misfit scenario F156H and Scenario F456H). All the unexpected responses had residuals of  $\pm 2$ . Again, the unexpected responses could be due to confusion with the survey instructions. Given the smaller magnitude

of residuals, the unexpected responses can be considered as reflecting fluctuations of degrees of difference.

The demographic characteristics of misfit participants indicated that they were comprised of varied gender and racial/ethnic identities. Participants also had varying degrees of teaching experience, ranging from less than one to three-to-five years, and were either attending or attended teacher preparation programs at either Boston College or other higher education institutions. Three out of the five misfit participants taught English Language Arts and Literacy at the elementary level. With regard to their social desirability summary scores, as measured by the 20-item Social Desirability Scale (Strahan & Gerbasi, 1972), these participants had total scores of between four and 16 (out of a total of 20), which were not particularly high compared to other participants in the sample.

Since these five participants had large weighted and unweighted mean square residuals (i.e., Infit- and Outfit-MNSQ > 2) and associated t statistics ( $ZSTD > 2$ ), I decided to delete these five participants and re-ran the analyses with 52 participants as suggested in the Rasch literature (Linacre, 2016; Smith, Linacre, & Smith, 2003). In addition, I removed the practice item in the final analyses. I discuss the psychometric properties of the final scale based on the 52 participants' responses, below.

***Results after deletion: Variable map.*** Figure 4.8 presents the variable map for the 15 scenarios and Figure 4.9 is the Andrich probability map. Similar to the results prior to the deletion (see Figure 4.6), the distribution of the revised scenarios mostly reflected what was expected. The location of Scenario F346H was lowered and the location of Scenario F126M was moved up as intended. Both revised scenarios filled the gap between Scenarios F236H & F346H and Scenarios F236M & F346M in the original distribution of the scenarios from the pilot study.

The revised Scenario F456L was located below Scenario F456M as expected. Moreover, though the location of Scenario F456L was right below Scenario F456M prior to the deletion, it pulled further away from Scenario F456M as intended. The only exception was Scenario F456H. This scenario became the second most difficult item after the deletion, but was expected to be the most difficult one with the earlier revision.

While the scenarios became more spread out and formulated a “ladder-like” continuum as intended, there remained some redundancy. Specifically, the two high-level Scenarios F156H and F456H had similar scale locations, and the same applied to the moderate-level Scenarios F236M & F346M and F156M & F456M. In addition, the four scenarios (F346L, F126L, F156L, and F236L) at the lower end of the distribution were not widely spread and could be redundant in the subsequent interpretation of persons’ scores.

Person separation was 2.08 (an increase from 1.76 prior to the deletion) with a person reliability of .81 (an increase from .76 prior to the deletion). The item separation was 8.74 (an increase from 7.51 in the pilot study and 5.70 prior to the deletion) with an item reliability of .99 (an increase from .97 prior to the deletion). This suggests that the deletion of poorly fitting cases resulted in participants’ ability to be more reliably differentiated. The revised scenarios after deleting the five misfit cases became more sufficiently and reasonably spread out to define the increasing levels of enactment of practice for equity compared to the distribution of the original scenarios.

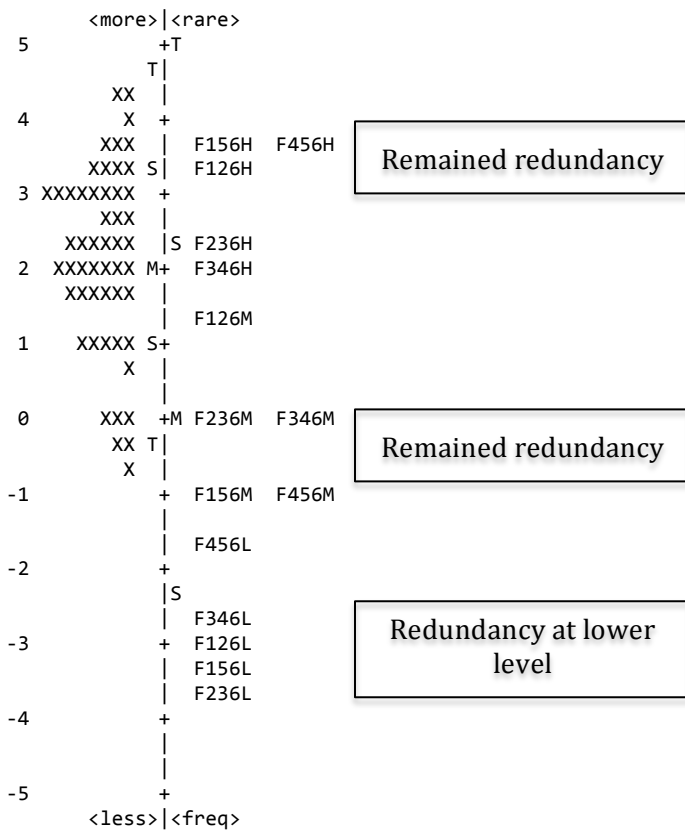
Lastly, comparing the locations of participants and items along the central line, the average practice level of participants was still higher than the average difficulty level of the scenarios and no participants were located next to the five low-level scenario items. This was similar to the results prior to the deletion. The Andrich map (see Figure 4.9) gives a more

precise overview of participant locations based on the probability that they would select certain response categories for each item. These results suggest that most participants were expected to choose *About the same* for the high-level scenarios and *Slightly higher* for the moderate-level scenarios. Most participants were expected to select *Much higher* to the low-level scenarios.

Figure 4.8

*Variable Map: Final Study (52 Cases)*

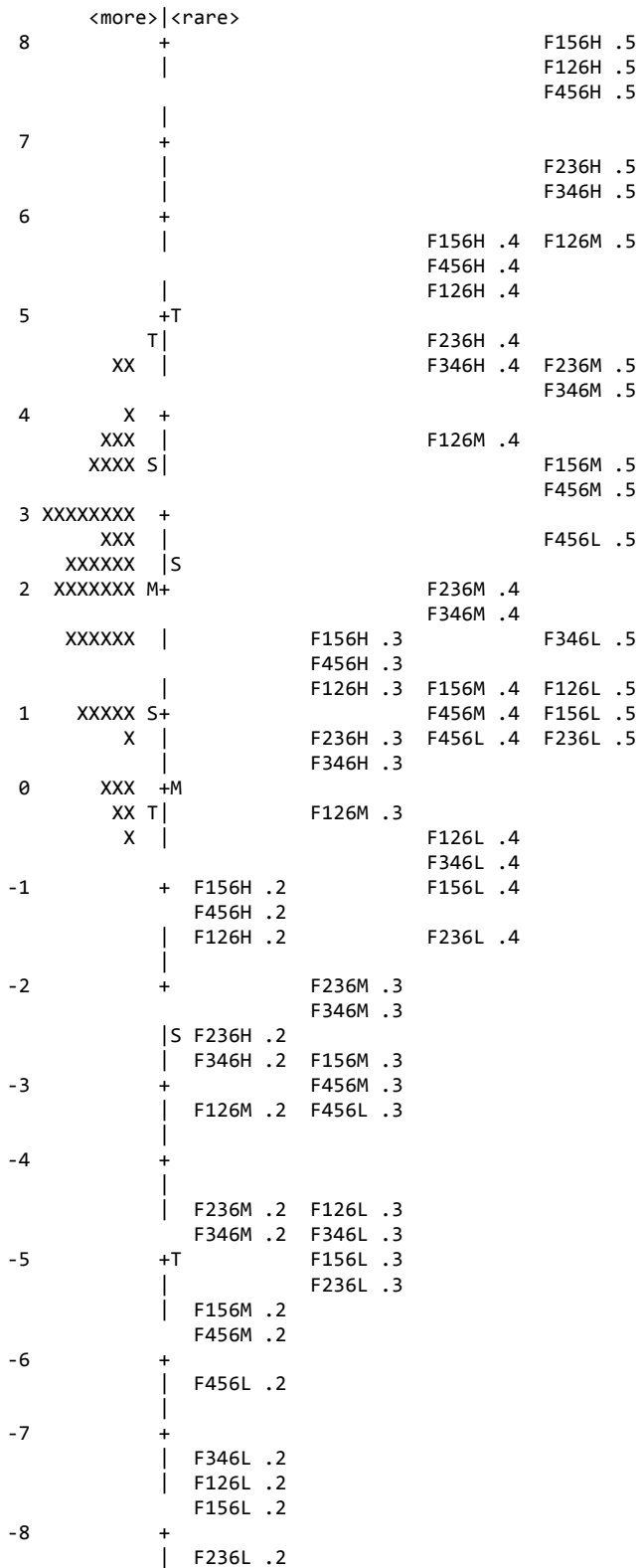
MEASURE Teacher - MAP - Scenario



As intended:  
Removal of the clusters,  
better spread scenarios,  
and towards a "ladder-  
like" continuum

Figure 4.9

*Cumulative Probability Map by Andrich Thresholds: Final Study (52 Cases)*



*50% Cumulative Probability Map by Andrich Thresholds: Final Study (52 Cases) (continued)*

-9            |  
              +  
             <less>|<freq>



The variable map in Figure 4.10 provides a scale interpretation based on participants' enactment of practice level. Using the scoring conversing table in WINSTEPS (see Appendix M on page 319 for the scoring conversing table), I identified the raw scores equivalent to the logit measures and presented the measure on the left of this figure in raw score units instead. In addition, I plotted horizontal lines that correspond to the average raw scores of 1, 2, 3, 4, or 5. This variable map provides a convenient and meaningful way to interpret the participant's level of enacting practice for equity along the scale based on the individual's sum score (see Table 4.15 for the Teaching Equity Enactment Scenario scale score interpretations).

For example, participants with raw scores in the range between 60 and 75 (logit measure between 3.35 and 10.24) were expected to select *Slightly higher* (score = 4) to the three high-level scenarios (Scenario F156H, F456H, and F126H) and *Much higher* (score = 5) to the scenarios below their location. According to the participants' self-reports about their classroom teaching, they are fully committed to their professional responsibilities (e.g., supporting the learning and life of diverse students, advocating on behalf of them, and contributing to the profession) and they reflect on their practice constantly. These participants hold and clearly communicate high expectations to all students, deliberately use a variety of sources (e.g., students' culture and prior knowledge) to design pedagogical strategies and instructional materials, and integrate assessment to scaffold learning and improve teaching.

For participants scoring in the 45 – 60 range (logit measure between 0.09 and 3.35), the participants were expected to choose *About the same* (score = 3) to the three scenarios capturing the moderate/high-level enactment (e.g., Scenarios F236H, F346H, F126M) and *Slightly higher* (score = 4) to the scenarios below their location. The participants generally care for and respect their students and hold high expectations to some groups of students. They involve students in

the process of classroom decision-making and encourage students to set criteria/goals for their learning and build understandings of their own. They also facilitate collaborative learning in the classroom, use some cultural examples in designing learning experiences, and provide constructive feedback to their students.

Participants scoring in the range of 40-45 (logit measure between -0.96 and 0.09) were expected to choose *Slightly lower* (score = 2) for moderate-level scenarios (e.g., Scenarios F236M and F346H) and *About the same* (score = 3) for the scenarios below their location. Their responses suggest that the participants generally care about and respect their students, but sometimes still engage in stereotypical thinking. They cooperate with some but not all parents, and sometimes involve their students in designing their learning experiences or setting learning goals. Their classrooms are generally a safe and inviting place where students sometimes interact to help each other learn and teachers occasionally provide feedback to students.

Figure 4.10

Construct Definition Map

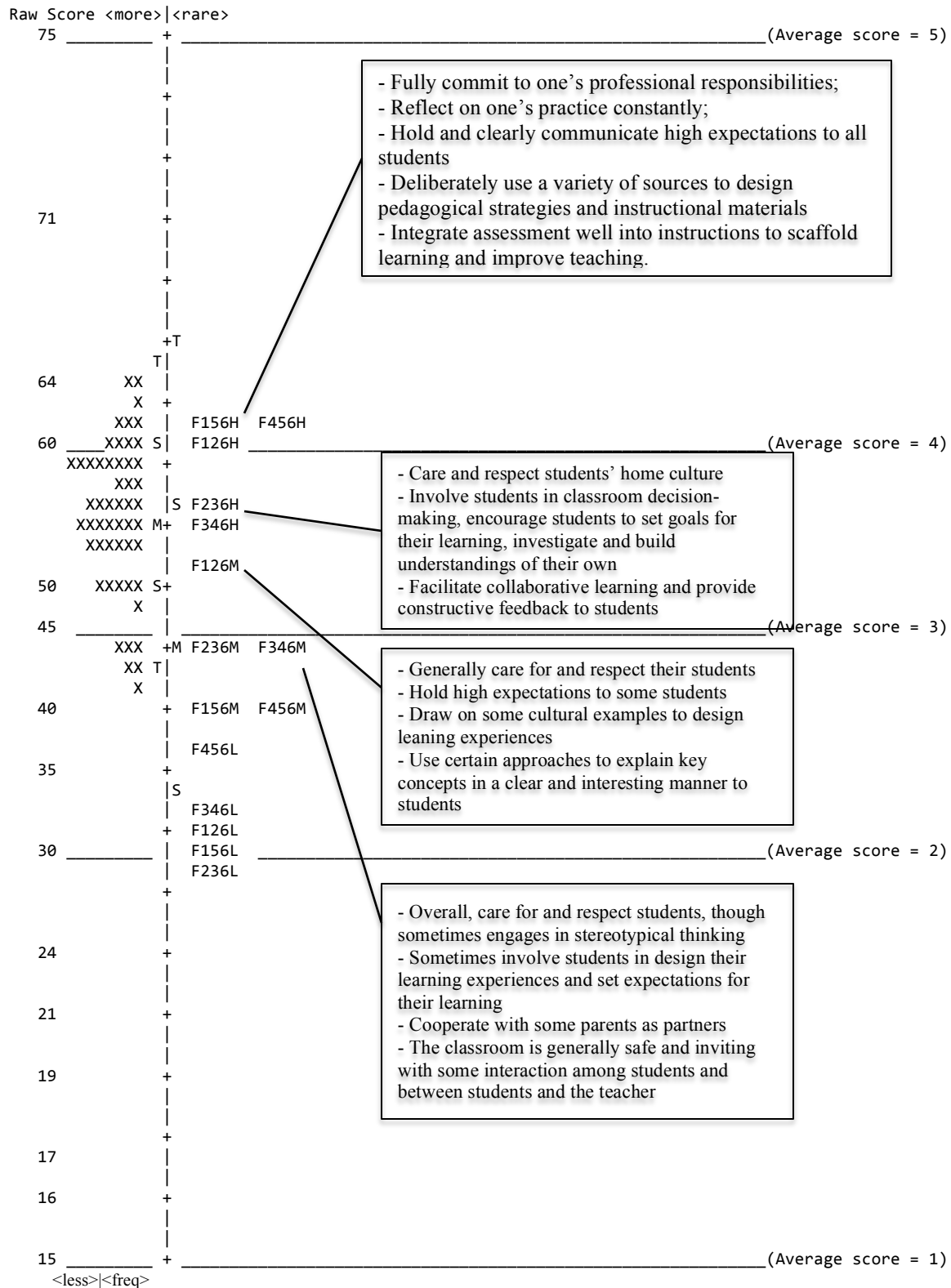


Table 4.15

*Teaching Equity Enactment Scenario Scale Score Interpretations*

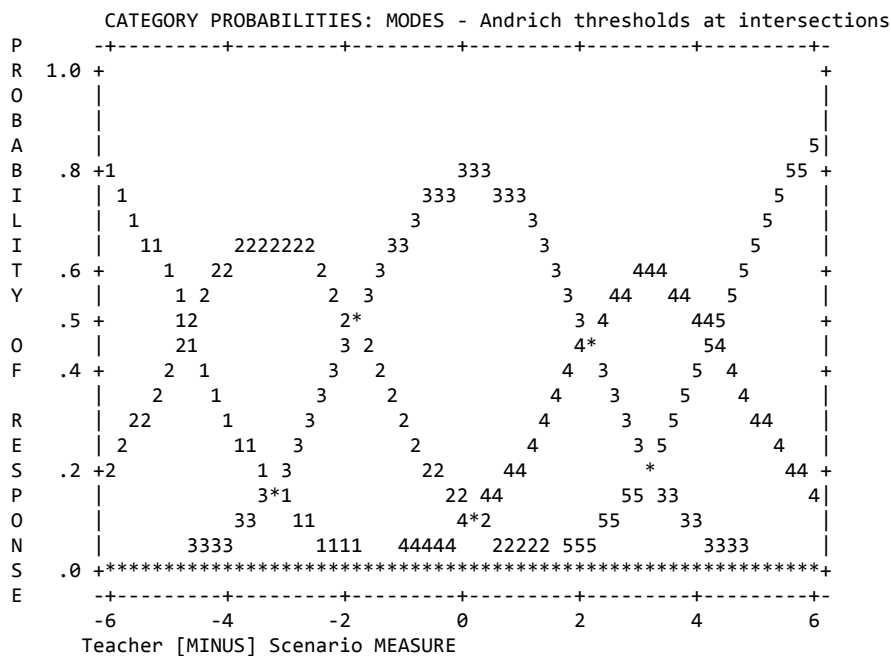
<b>Scenario Score</b>	<b>Enactment Level</b>	<b>Description of score</b>	<b>Example of Scenario</b>
75	Extremely high level of enactment	You are “much higher” than all scenarios presented	
60 - 74	Very high level of enactment	On average, you are “slightly higher” than Scenario F156H and “much higher” than the scenarios below this section	<b>F156H</b> (Facet score = 9) Megan fully embraces her responsibility to identify and challenge classroom and school practices that promote inequities for students. Megan sets cognitively challenging goals and communicates to her students clearly and consistently. She purposefully draws upon a variety of sources to cultivate their conceptual understanding and encourages students to challenge information in textbooks. She deliberately uses various pedagogical strategies to capture students’ interests. Megan also works with others in a professional community to pose questions, reflect on her own assumptions, and proactively respond to student needs.
55 - 59	High level of enactment	On average, you are “about the same” as Scenario F346H and “slightly higher” than the scenarios below this section	<b>F346H</b> (Facet score = 8) Katherine cares for and respects her students. She encourages students to be independent learners and to investigate and build understandings of their own, and she involves them in setting criteria and goals for their learning. She constructs her teaching practice to be engaging to all students, and integrates a variety of assessment approaches into her teaching. Katherine interacts with students to provide constructive feedback and adjusts her practice appropriately. She monitors and facilitates collaborative learning among her students.
45 - 54	Moderately high level of enactment	On average, you are “about the same” as Scenario F126M and	<b>F126M</b> (Facet score = 7) Tim holds high expectations for some students in his class and mostly communicates these expectations clearly. He generally sees students’ home culture as a strength and collaborates with some parents/caregivers. He sometimes lets his

		“slightly higher” than the scenarios below this section	students choose a topic consistent with their interests to further their learning. Tim sometimes draws on cultural examples to design learning experiences that are relevant to students. He utilizes a selected number of approaches to explain key concepts. His explanations are clear and interesting to all students.
40 - 44	Moderate level of enactment	On average, you are “slightly lower” than Scenario F236M and “about the same” than the scenarios below this section	<b>F236M</b> (Facet score = 6) Tracey cooperates with some parents/community members and draws on some students’ culture as examples to design their learning experiences. Overall, she genuinely cares for and respects her students, though sometimes engages in stereotypical thinking. Tracey sometimes involves her students in designing a lesson or setting classroom rules. Although she often has students concentrate on their own work, she sometimes encourages collaboration among students. Tracey’s classroom is inviting and safe for some students.
30 - 39	Low level of enactment	On average, you are “slightly lower” than Scenario F456L and “about the same” than the scenarios below this section	<b>F456L</b> (Facet score = 3) Dave believes good teaching is executing a set of techniques to ensure that students attain curriculum expectations for their grade/year levels. He solely relies on standards or curriculum documents to identify learning priorities and teaching approaches. He designs assessments on his own and generally uses them to check whether students meet the minimum academic standards. Dave reviews his students’ test results, sometimes altering his practice to boost their scores.
15 - 29	Extremely low level of enactment	On average, you are “much lower” than F236L and all other scenarios presented	<b>F236L</b> (Facet score = 3) Christine occasionally engages with parents but generally she sees this as unnecessary. She has a quiet, reserved manner with her students, approaching all students the same way. Christine, rather than involving the students, makes most of the decisions in the classroom. Because she believes that students should work individually, Christine does not see the need to facilitate interactive skills and usually assigns work for students to carry out on their own. Her classroom is arranged to be a generally quiet, non-interactive space.

**Results after deletion: Categorical characteristic curves (CCCs).** Figure 4.11 below outlines the categorical characteristic curves which represent the probability of responding to each of five categories for any participant on any scenario item. The intersection where any curve meets is the Andrich threshold estimates  $\tau_j$  which indicate the probability of moving from one category to the next. The Andrich thresholds as reported in WINSTEPS (Table 3.2 – not shown) were -4.65 ( $\tau_1$ ), -1.83 ( $\tau_2$ ), 2.20 ( $\tau_3$ ), and 4.28 ( $\tau_4$ ). The threshold estimates were in the intended ordered pattern, and each of the five response categories had a dominant area of distribution that covered a wide range. Each category had about the same probability in height, though the third category (i.e., *about the same*) was slightly higher. Overall, the response categories functioned as intended and all response categories were used by participants. This ordered pattern and the roughly equal probability of each category was consistent in the pilot study, the final administration prior to the deletion, and this last analysis after the deletion.

Figure 4.11

*Categorical Characteristic Curves (CCCs): Final Study (52 Cases)*



***Results after deletion: Goodness-of-fit.*** Table 4.16 below presents item misfit information after deleting the five extremely misfit cases. Using the liberal threshold of 1.3 for both Infit- and Outfit-MNSQ, this left two remaining misfit items. The first misfit item was Scenario F346L with an Infit-MNSQ of 1.79 (Infit-ZSTD = 3.2) and an Outfit-MNSQ of 1.47 (Outfit-ZSTD = 1.8). The second misfit item was Scenario F126H with an Infit-MNSQ of 1.74 (ZSTD = 2.9) and an Outfit-MNSQ of 1.69 (Outfit-ZSTD = 2.7). The deletion of the five extremely misfit cases after careful consideration and measurement justification resulted in fewer numbers of misfit items, as well as a smaller magnitude of weighted and unweighted mean squares and associated t statistics. Specifically, Scenarios F126L, F156H, and F236L were no longer misfit items after the deletion. Most importantly, Scenario F156L was no longer a misfit item, but had been a misfit item in both the pilot study and prior to the deletion. This suggests that the specific revision done on Scenario F156L after the pilot study might have addressed the issue as expected. Again while Scenario F346L remained a misfit item and Scenario F126H became a misfit item, the magnitude of mean squares and t statistics was within a reasonable range (i.e., ZSTD < 3.0).

Table 4.16

*Item Misfit: Final Study with 52 Participants*

Scenario	Entry Number	Item Difficulty	Infit		Outfit	
			MNSQ	ZSTD	MNSQ	ZSTD
<b>F346L</b>	<b>14</b>	<b>-2.70</b>	<b>1.79</b>	<b>3.2</b>	<b>1.47</b>	<b>1.8</b>
<b>F126H</b>	<b>11</b>	<b>3.23</b>	<b>1.74</b>	<b>2.9</b>	<b>1.69</b>	<b>2.7</b>
F456L	5	-1.58	1.40	2.1	1.32	1.7
F156H	8	3.60	1.29	1.4	1.21	1.0
F126M	1	1.34	0.94	-0.2	0.94	-0.2
F456H	13	3.52	0.93	-0.3	0.90	-0.4
F346M	7	-0.08	0.80	-1.1	0.90	-0.5
F156M	15	-0.93	0.82	-1.0	0.86	-0.7
F346H	2	2.10	0.77	-1.0	0.75	-1.1
F156L	3	-3.18	0.67	-1.6	0.71	-0.9
F236H	4	2.46	0.71	-1.3	0.70	-1.3
F126L	6	-2.93	0.70	-1.5	0.64	-1.4
F456M	10	-1.09	0.70	-1.8	0.70	-1.7
F236L	9	-3.66	0.69	-1.3	0.52	-1.3
F236M	12	-0.08	0.66	-1.9	0.67	-1.9

Table 4.17 below presents the person misfit information for six cases after the initial deletion. The following discussion of misfit cases focuses on the first five participants with both Infit- and Outfit-MNSQ and ZSTD larger than 2. After deleting the five extremely misfit cases, there were still four participants with large MNSQ and ZSTD of 2 and 3. As expected, the deletion resulted in a lesser number of misfit cases and much smaller magnitude of fit indices (MNSQ and ZSTD had a range between 2.0 and 8.0 before the deletion).

The first misfit case #17 was a high-scoring respondent with the enactment measure of 2.37, who gave a lower than expected response to Scenario F346L, a low-level misfit item, with a large standardized residual of -5. The unexpected response to Scenario F346L might be due to carelessness. Specifically, it was possible that the teaching practice described in Scenario F346L was so unfavorable (and particularly so after seeing four moderate or high-level scenarios in a



row) that this participant automatically gave the *Slightly lower* response without thinking carefully back to the survey instructions. In fact, the same rationale could apply to two other misfit participants (participants #3 and #27), who also gave lower than expected responses to Scenario F346L.

Participant #16 was a low-scoring respondent with an enactment measure of -0.11. This person gave a higher than expected response to Scenario F126H (standardized residual 3), a high-level misfit item. This participant also gave a higher than expected response to another high-level scenario F156H with standardized residual of 2. It is possible that the higher than expected responses to high-level scenarios could be due to the effect of social desirability. This participant had a social desirability score of 14 out of 20. In fact, similar patterns of the same two scenarios were seen in participant #17 (a social desirability score of 17 out of 20) who tended to give even higher than expected responses to high-level scenarios with a standardized residual of 2. Details about the effect of social desirability are discussed in the following section.

Participant #3, a low-scoring respondent, gave a lower than expected response to the low-level misfit item Scenario F346L and a higher than expected response to the high-level misfit item Scenario F126H. Looking closely into the all other responses of this participant, this participant gave *About the same* response to 12 items in a row and *Slightly lower* response to the last three items. It was evident that this participant likely did not read the scenarios and tried to get through the survey as quickly as possible.

Participant #27, an average enactment level participant (ability measure of 0.09), gave a lower than expected response to low-level misfit item F346L with a standardized residual of -4. As discussed earlier, the reason for the unexpected response to this item may be due to the entry

order of this low-level scenario. This was also the only item for which this participant had an unexpected response with a large standardized residual.

The last misfit participant (#12) had an enactment measure of 0.90. This participant had a higher than expected response to the high-level misfit item F126H with a standardized residual of 2 and a lower than expected response to the low-level misfit item F346L with a standardized residual of -2. In addition, this participant selected a lower than expected response to a high-level item F156H with a standardized residual of -2. Based on the magnitude of the misfit and this participant’s responses to other items, it seemed that the unexpected responses were a matter of degree.

Table 4.17

*Person Misfit: Final Study with 52 Participants*

Participants	Person Ability	Infit		Outfit	
		MNSQ	ZSTD	MNSQ	ZTSD
17	2.37	3.17	3.9	3.41	3.9
16	-0.11	2.93	3.7	2.89	3.5
27	0.09	2.38	2.9	2.20	2.5
3	-0.53	2.19	2.6	2.09	2.3
12	0.90	2.03	2.5	1.94	2.2
13	1.51	1.86	2.1	1.66	1.7

***Results after deletion: Effect of Social desirability.*** To understand the effect of social desirability on item responses, Pearson correlations between social desirability scores and responses to each item were obtained. Since higher scores (maximum score was 20) suggest higher levels of social desirability, ideally correlations should be low (i.e., less than 0.3). If an item has a significant and high correlation (i.e., above .5) with the social desirability scores, the item may be removed or revised.

The remaining 52 participants had a mean social desirability score of 13.92 ( $SD = 3.29$ ) ranging between seven and 19. Pearson correlations between the social desirability sum scores and responses to each item were obtained and are presented in Table 4.18 below. All items had correlations less than 0.3, except for Scenario F156H, which captures high-level enactment of practice for equity. Responses to Scenario F456H were also positively correlated with the social desirability scores (A significant Pearson correlation of .261). Interestingly, scenarios that had higher correlations (though not necessary significant) tended to be the ones capturing high-level enactment such as Scenarios F126H (a misfit item) and F236H. This suggests that responses to the high-level scenarios might be more prone to the effect of social desirability. One plausible explanation to this pattern could be that the teaching practice captured in these high-level scenarios was favorable, and participants would believe this is what their practice looks like in the classroom. Because my earlier concern during the instrument development stage was about the effect of social desirability on the low-level scenarios, I took multiple revisions to minimize the threat, and as a result, it seems that responses to low-level scenarios are not particularly affected by social desirability.

Table 4.18

*Pearson Correlations: SDS Sum scores and Scenarios*

<b>Scenarios</b> (order of items based on the entry number)	<b>Correlations with Sum of SDS</b>
F126M	.199
F346H	-.002
F156L	.091
F236H	.167
F456L	-.101
F126L	.091
F346M	.079
F156H	.345*
F236L	.112
F456M	.150
F126H	.178
F236M	-.073
F456H	.261*
F346L	.040
F156M	-.018

\* Correlation is significant at the 0.05 level (one-tailed)

**Residual analysis.** As stated in Chapter Three, the purpose of residual analysis is to check for non-random factors left unexplained in the data as well as the remaining structure of standardized residuals after extracting variation accounted for by the Rasch model. To conduct residual analysis, I obtained the standardized residuals in WINSTEPS. I then performed Factor Analyses (FA) with Principal Axis Factoring method in SPSS to the standardized residuals. The results showed that the determinant was .006. The non-zero determinant suggests that there existed some variability in the matrix to allow for factor analysis. This is less preferable to see for the residual analysis. The KMO was .276, which indicates that the shared variances among items were small. A KMO of .276 that is small and closer to zero is desirable for the purpose of residual analysis. The Barlett's test was still significant and some correlations remained among

the residuals. This was less than ideal, as a non-significant Barlett's test, which indicates a diagonal matrix for the standardized residuals, is preferable.

Table 4.19 presents the total variance explained by factors. The first factor had an eigenvalue of 3.04 and explained 20.27% variance in the data. The first factor was followed by six factors with eigenvalues larger than one. There remained one clear factor, as outlined on the Scree plot (see Figure 4.12). This suggests that there was still an unexplained non-random dimension among the standardized residuals after extracting variance accounted for by the Rasch rating scale model.

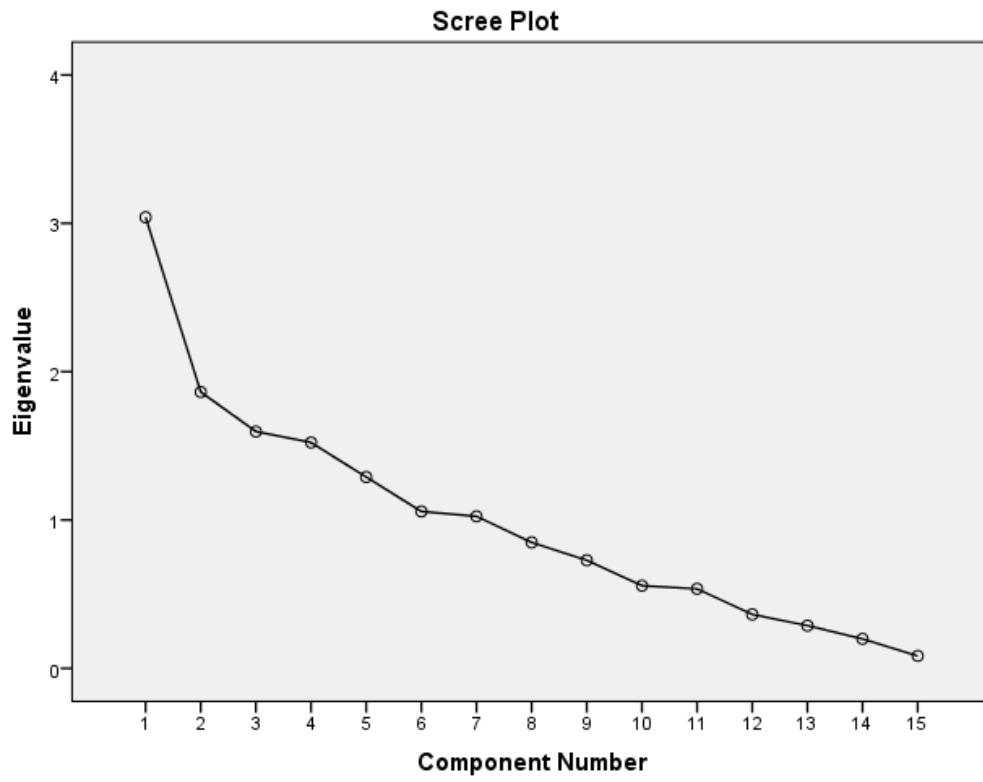
Table 4.19

*Total Variance Explained: Standardized Residuals*

<b>Factor</b>	<b>Initial Eigenvalues</b>		
	<b>Total</b>	<b>% of Variance</b>	<b>Cumulative %</b>
1	3.041	20.273	20.273
2	1.863	12.419	32.693
3	1.596	10.641	43.333
4	1.523	10.152	53.485
5	1.290	8.599	62.083
6	1.058	7.051	69.134
7	1.025	6.831	75.965
8	0.849	5.659	81.624
9	0.729	4.858	86.482
10	0.557	3.716	90.198
11	0.536	3.575	93.773
12	0.364	2.424	96.197
13	0.288	1.921	98.118
14	0.199	1.324	99.442
15	0.084	0.588	100.000

Figure 4.12

*Scree Plot: Standardized Residuals*



## Summary of Results

In this chapter, I discussed the results of the Rasch analyses on both the pilot study and the final full-scale administration. Using the Rasch rating scale model, the results of the pilot study suggested that the locations of the 15 scenario-style items captured the three distinct levels of enactment of practice for equity from the low to the high levels as expected. Given that the *proof of concept* was achieved, the targeted revisions were done to address three misfit items and to obtain a more reasonably spread out scale. The results of the final administration and the careful elimination of five misfit persons produced a scale where the scenarios are sufficiently spread to define the increasing levels of the unidimensional construct, teachers' enactment of practice for equity, as intended. Results regarding the distribution of the scenario-style items, the separation indices, and the reliability estimates indicate that the Rasch-based scenario-style scale can measure teachers' self-reports about their enactment of practice for equity in a reliable and meaningful way. In addition, the overall results of goodness-of-fit, CCCs, and residual analysis suggest that the data fit the Rasch rating scale model and support the *a priori* theory of the intended distribution of the scenarios and that the scale is consistent with Rasch measurement principles.

In the final chapter, I discuss the findings of this study, its contributions to the existing body of research on developing instruments to measure aspects of equity-centered social justice oriented teaching (e.g., beliefs, self-efficacy, knowledge, and practice), and the implications for research on teaching and the field of measurement. I also present the limitations of the study and recommendations for future research with a focus on the efforts needed to improve the validity, utility, and feasibility of this scale.

## CHAPTER FIVE: DISCUSSION

In the previous chapter, the empirical findings confirm that a Rasch-based scenario-style scale measures a hierarchical continuum defining a unidimensional construct corresponding to teachers' self-reports about their enactment of practice for equity, in a reliable, valid, and meaningful way. In this chapter, I discuss this finding and this study's contributions to the existing body of research on developing instruments to measure aspects of equity-centered social justice oriented teaching (e.g., beliefs, self-efficacy, knowledge, and practice) and the field of measurement. This is followed by discussion of the limitation of the study, recommendations for future research, and implications and conclusion.

### Overview of Findings

The central focus of this study was reflected in the research question, "*Can the construct of teachers' self-reports about their enactment of practice for equity be measured reliably and meaningfully by using a Rasch-based scenario-style scale?*" To answer this question, this study employed a novel approach of combining Rasch measurement theory (Rasch, 1960/1980; Wright & Masters, 1982) and Guttman's facet theory (Borg & Shye, 1995; Guttman & Greenbaum, 1998) to develop 15 scenario-style Likert-type items. The Rasch rating scale model was used to analyze the pilot study data of 73 participants and the final full-scale administration data of 52 participants (after deleting the five poorly fitting respondents). The results of the pilot study demonstrated the *proof of concept*; that is, the locations of the scenarios capture the low, the moderate, and the high level of teachers' self-reported enactment of practice for equity as intended. In other words, the empirical data confirmed the hypothesis: the high-level scenarios were more difficult for participants to select higher ratings (i.e., *About the same* or higher) than the moderate- and low-level scenarios, and the moderate-level scenarios were more difficult for



participants to choose higher ratings than the low-level scenarios. Moreover, the hierarchical order of scenarios was very similar between the New Zealand and the U.S. participants as well as between novice and experienced participants – evidence for *measurement invariance*. The fit analysis and the think-aloud exercise were used to aid the item revisions, which were intended to obtain a scale that defines increasing levels of enacting teaching practice for equity with more sufficient spread and separation among scenarios.

Based on the results of the final full-scale administration, the final outcome of the study is a scale of 15 scenarios that can be used to measure teachers' self-reports about their levels of enactment of practice for equity in the classroom in a reliable, valid, and meaningful manner (See Figure 4.10 on page 143 for the construct definition map and Table 4.15 on pages 144 and 145 for the interpretation of raw scale scores). The findings suggest that the empirical data from both research phases fit the Rasch model. This means that the scale confirms the *a priori* theory derived from the theoretical base of the construct and the requirement of Rasch measurement principles, including unidimensionality, variation and a hierarchical order of the items, as well as a uniform continuum of the construct.

### **Discussion of Findings**

This dissertation contributes to the existing body of instrument development studies intended to measure aspects of equity-centered social justice teaching in several ways: a) the scenario-style scale developed in this study measures the perception of enactment of teaching practice for equity rather than proxies of teaching practice, b) the instrument puts equity 'front and center' in a way that each of the 15 scenarios includes the practice of Facet 6 – "recognizing and challenging inequities," c) the novel approach of combining Rasch measurement theory and Guttman's facet theory to develop a scenario-style scale appears to be successful in measuring a

complex construct, and d) the use of the Rasch-based scenario-style scale allows meaningful interpretations of one's levels of enacting teaching practice for equity, measuring growth, and making comparisons between groups.

Most existing instruments that intend to capture the theoretical construct similar to this study instead measure teachers' knowledge, attitudes, beliefs, awareness, and self-efficacy of enacting equity-centered social justice teaching. While these factors are important predictors of teachers' practice in the classrooms, scholars (e.g., Sleeter, 2001) have been calling for research in teacher education to measure teachers' practice. Although there exist validated instruments that measure what teachers are able to do in the classroom, these instruments are not grounded in the theories of equity-centered social justice teaching. While it should be acknowledged that this instrument is currently a self-report type of instrument, with further revision this instrument could be used as an observational tool.

In addition, although most other instruments reviewed in this study measure aspects related to equity-centered social justice teaching, close examination reveals that the political and critical aspects of enacting equity-centered social justice teaching (i.e., recognizing and challenging inequities, advocating on behalf of students) are often diluted in the theoretical foundations, if not left out completely. As a result, most instruments reviewed do not appear to capture, for example, beliefs or attitudes towards recognizing and challenging inequity or advocating on behalf of students for structural change. The lack of focus on equity-centered teaching practice is also applied to the body of instruments that measure what teachers are able to do in the classroom.

This instrument contributes to the existing body of studies by emphasizing the practice of recognizing and challenging equity (Facet six) as the core of teaching practice for equity in the

process of construct mapping and item development. This was achieved by making a clear connection between the theoretical framework of equity-centered social justice teaching, the syntheses/programs of research chosen to provide the operational definition of teaching practice for equity, and the decisions about the scenario structure. Specifically, the theoretical framework of *equity-centered social justice teaching* informed the selection of five international syntheses and programs of research, which provided information to define the construct of teaching practice for equity (Cochran-Smith et al., 2016). Following an iterative process of reviewing the syntheses/programs of research and construct clarification with content experts, the construct of teaching practice for equity included six main characteristics (i.e., six facets) with one of the six facets named “recognizing and challenging inequities.” The decision that this facet must be a stand-alone facet, rather than implied or embedded with other characteristics, reflects the theoretical framework of equity-centered social justice teaching. Given the importance of the chosen theoretical frameworks and construct definition of teaching practice for equity, the scenario structure was made to ensure that all 15 scenarios capture the practice of recognizing and challenging inequities.

This study also contributes to the existing body of instruments measuring aspects of equity-centered social justice teaching and the field of measurement by using a novel approach of combining Rasch measurement theory and facet theory to develop a scenario-style scale. All presently existing instruments measuring teaching practice and aspects of equity-centered social justice teaching use simple, one-statement Likert-type questions. While that type of instrument can be useful (and is easier to construct), the simple statement Likert-type items do not authentically capture the complex nature of teaching with a commitment to equity and social justice. If the nature of teaching for equity and social justice is inherently multi-faceted and

multi-barreled, then an instrument should attempt to capture the complexity. Applying Rasch measurement theory and Guttman's facet theory, the 15 scenarios were developed to measure the increasing levels of enactment of practice for equity (i.e., from the low, to the moderate, and to the high level) as a unidimensional ladder-like construct. The empirical data confirmed the intended hierarchical order and the fit analyses suggested that the scale fulfilled the requirement of the Rasch measurement principles. The scenario-style scale appears to measure the complexity of enacting practice for equity based on teachers' self-reports in a reliable and valid manner.

Given that the six interconnected facets define the construct of practice for equity and each facet includes a set of teaching characteristics, enacting equity-centered teaching practice at different levels necessarily involves a combination of practice related to all six facets. For example, at the lower range of the continuum, a teacher is characterized as acknowledging the importance of improving outcomes for all students; however, he/she is unaware of the impact of his/her practice and cultural positioning and tends to attribute the educational underachievement to the deficits in learners and their families. The teacher holds high expectations for very few students and rarely sets and communicates cognitively challenging and worthwhile learning goals. The teacher designs the learning opportunities that are rarely relevant to students' lives, cultural experience, the curriculum, and the valued learning outcomes. The relationships between the teacher and the students are rigid, and the teacher seldom recognizes students' home culture and connects his/her instructional practice to students' cultural knowledge. The teacher rarely facilitates collaboration among students or shares the decision-making power and responsibility with students. The teacher does not integrate the assessment well into the

instructional activities to help students learn, provide timely feedback, and adjust their instruction. He/she tends to work alone and rely on familiar instructional approaches.

Moving up the construct of practice for equity, a teacher may be characterized as recognizing his/her responsibility to enhance all (and marginalized) students' learning outcomes. The teacher does not accept the deficit thinking to explain educational under-achievement of historically marginalized students and sometimes engages in critical reflection of his/her practice. The teacher holds high expectations for some groups of students and mostly communicates clearly with minor inconsistencies. The teacher designs the learning opportunities that are mostly relevant to students, but sometimes the learning opportunities are not relevant to students' lives, the curriculum, and the valued outcomes. Overall, he/she cares about and respects the students, but sometimes still engage in stereotypical thinking. The teacher cooperates with some but not all parents, and sometimes embeds students' cultural knowledge/experiences into learning experiences. The classroom is a safe and inviting place where students sometimes interact to help each other learn, and the teacher sometimes involves the students in designing their learning experiences or setting learning goals. The classroom assessment is mostly integrated into the instructional activities, and the teacher occasionally circles and provides feedback to students. The teacher sometimes works with others to engage in investigating the impact of their practice on student learning, but mostly relies on certain kinds of sources to identify learning priorities for students.

Finally, at the upper lever of the construct, a teacher may be characterized as deeply committed to their professional responsibilities to enhance all (and especially marginalized) students' learning outcome. The teacher explicitly rejects deficit thinking of historically marginalized students through critical self-reflection, sharing classroom decision-making power

with students, and advocating structural change on behalf of students. The teacher holds and clearly communicates cognitively challenging learning goals for all students and deliberately uses a variety of sources and instructional materials to design pedagogical strategies. The teacher genuinely cares and respects students' home culture, and makes students' learning experience relevant and meaningful to their lives. The teacher effectively builds a learning community through facilitating collaboration among students and fostering an inclusive learning environment. The teacher also integrates various assessment approaches well into instructional practice to scaffold learning and improve teaching. He/she collaborates closely with others to engage in inquiry-based research on practice, proactively reflects on his/her own practices, knowledge, biases and assumptions, and consistently takes action to support student learning outcomes even though it might not resonate with his/her existing practice or beliefs.

Given that the empirical data fit the Rasch rating scale model, the calibrated scenarios allow meaningful interpretation of one's enactment level along the scale. As presented in Figure 4.10 (page 143) and Table 4.15 (pages 144-145) a respondent's raw score along the scale has a corresponding description of this participant's current level of enacting practice for equity. The score descriptors along the scale provide the diagnosis analysis of an individual's teaching practice, a unique benefit for instruments using Rasch measurement theory.

The use of Rasch measurement theory in developing the Teaching Equity Enactment Scenario Scale provides several advantages that are absent among instruments using the Classical Test Theory approach, which is the main approach utilized by the studies reviewed in Chapter Two. First, using the Rasch measurement theory, the scenario-style scale was developed according to the Rasch measurement principles and theoretical construct of teaching practice for equity. Using the Rasch rating scale model, a confirmatory statistical model, the calibrated

scenarios confirm the hypothesized order and fit the Rasch model. In this case, the resulting variable maps presented in Chapter Four provided the evidence on construct validity for this instrument.

Additionally, based on the Rasch rating scale probabilistic model, both respondents' and items' ordinal-level raw scores are transformed to the interval level measurement unit in *logits*. When respondents' enactment ability measures and item difficulty estimates are placed on the same interval-level scale, the difference between person measures and item estimates yields the probability of one's response to an item. Equal interval, under the Rasch framework, means that, for example, one logit difference between a person measure and an item estimate gives the same probability of choosing a specific response (e.g., *Slightly higher*) versus others (e.g., *About the same* and lower ratings) regardless of the locations along the scale. However, this is not necessary the case for instruments using the CTT approach, which relies on raw scores, since responses are often collected by using a Likert-type response format. The problem with raw score is that ten-point person raw score differences can be interpreted differently depending on where the difference is located in the distribution of raw scores.

Another related advantage of using Rasch measurement framework is the reproducibility of a Rasch-based scale. As outlined in Chapter Two, the estimation of item parameters and the estimation of person parameters can be separated from one another in the joint probability function of Rasch models (Rasch, 1966). Item estimates can be obtained by using the total item raw scores without dealing with the person parameters, and person measures can be obtained by knowing the total person raw scores without dealing with item parameters (Rasch, 1966). The *inferential separability* and *sufficiency* features of Rasch probabilistic models provide necessary conditions for having *objective measurement* (Rasch, 1966; Wright, 1967; Wright & Masters,

1982) that enables objective comparison. This suggests that the Rasch-based Teaching Equity Enactment Scenario scale is not dependent on the responses of sampled participants in this study, and participants' measures along the scale are not dependent on the scenarios that formulate it. In other words, item difficulty estimates and participants' enactment ability measures are not arbitrary. The *objectivity* or *invariance* of the Teaching Equity Enactment Scenario scale was demonstrated through variable maps obtained in different phases of the study, as well as comparisons between subgroups conducted in the pilot. The scenario scale can be used to diagnose enactment ability of persons along the scale, provide meaningful interpretation of person location in relation to the construct, and measure differences across groups and change over time (Hambleton & Jones, 1993; Wright & Masters, 1982). This is not the case for instruments using the CTT approach, in which item estimates are sample-dependent (thus the scale is not *fixed*) and person measures are test-dependent (thus the measures are arbitrary and depends on the difficulty of the items).

### **Limitations**

While the results of this study are encouraging and promising, this study has several limitations. The primary limitations relate to the small sample size and representativeness of the sampled participants. The last chapter presented the sample size of 73 for the pilot study and 52 for the final full-scale administration. While a large sample size was not required for conducting the Rasch analyses, the sample sizes of both phases were lower than the desirable number (between 75 and 150 - five to ten times of the scenarios) recommended by the general guideline for item analysis. The convenience sample of this study recruited through personal and professional networks was not representative of the larger population of K-12 teachers and teacher candidates who were prepared by teacher preparation programs within high education



institutions. In addition, to my best knowledge, no participants in this study were prepared through alternative routes or enter the profession through alternative certifications. However, this study successfully recruited a diverse pool of participants from different countries and cultural backgrounds (i.e., New Zealand and the U.S. participants), having varying years of teaching experience, and attended programs other than at Boston College. The participants also reported teaching at different levels of K-12 education and different academic subject areas. Therefore, while the sampled participants were not representative of the target participants, this study's sample still reflected a diverse sample of teachers and teacher candidates.

In addition, the survey completion rate (approximately 60%) was less than ideal for both the pilot study and the final full-scale administration. The high dropout rate was not surprising and was likely due to both the length of the scenarios and their multi-barreled nature. Moreover, unlike typical survey instructions for Likert-type items that ask participants to give a rating about a simple statement, the survey instructions in this study asked respondents to reflect on their practice, compare their practice against the practice described in a scenario, and give a rating. This task was undoubtedly atypical and even overwhelming to some participants. The less than ideal completion rate suggests that respondents may have felt frustrated, confused, or overwhelmed, which contributed to sources of measurement error. The low completion rate also has influence on the feasibility of this instrument. If preparation programs are to use this instrument to measure growth of their teacher candidates from entry to their first three years into the profession, this current instrument could be potentially burdensome rather than useful.

The multi-barreled nature of scenarios does not fall squarely within the convention of item development, and has been of concern since the beginning of the instrument development. Specifically, each scenario consists of approximately five sentences, which capture the practice

of three of six facets. While a scenario-style item like this aligns with the theoretical framework of teaching for equity and justice that views teaching as complex, it also raises the question of how participants respond to a scenario constructed of several statements. During the instrument development process as outlined in Chapter Three, an extensive effort was taken to make each scenario more holistic, engaging, and story-like (versus rigid and fragmented statements that are assembled together). Also, survey instructions were presented to guide participants to “read the scenario holistically.” However, the feedback from two key informants and an informal feedback received from a dropout respondent in the pilot phase indicated that some participants were caught in the specific sentences of a scenario, and were unsure how they should respond. Also, the first and the last sentence in a scenario seemed to have more weight than sentences in the middle of a scenario and tended to have more influence on participants’ overall impression of the practice in a scenario. The multi-barreled nature of scenario-style items may also contribute to the low completion rate as discussed above.

Another limitation of this study is that the instrument is currently a self-report instrument. While the scenarios are intended to measure the enactment of teaching practice for equity, the instrument is for teachers or teacher candidates to self-report what their teaching practice is like comparing to the practice described in a scenario. In other words, the instrument product of this study is not an observational tool to capture teachers’ practice in the classroom. Plus, due to the nature of the construct (enactment of teaching practice for equity), it is more favorable to consider one’s own practice to be at the higher rather than the lower levels. Therefore, participants’ responses based on their self-reports can be biased by factors such as social desirability, which contributes to systematic measurement error, as discussed in the previous chapter. Another related limitation with the self-report instrument is that when respondents are

asked to reflect on their own practice and compare against practice in a scenario, it is unclear whether they reflect on what they know the teaching practice should be or what their teaching practice really looks like in the classroom.

The final limitation of this study has to do with the design. The researcher was not able to contact the survey respondents to understand their thought process especially related to their unexpected responses to misfit items. This limited the opportunities to discover issues or flaws in the scenario scale and to identify solutions for item revisions. To make up this limitation, the think-aloud exercise was conducted with three key informants, who were (former) teachers and/or measurement experts, to help identify potential issues with the misfit items and brainstorm best possible solutions. While the think-aloud exercise with the three key informants could not equate to the kind of opportunities to talk to the survey participants; nevertheless, the feedback was useful.

### **Recommendations for Future Research**

The conclusion of this instrument study offers a promising foundation to providing empirical evidence about the extent to which teachers or teacher candidates enact practice for equity in the classroom – an essential outcome of teacher preparation. This empirical evidence can contribute to research that seeks to improve programs and/or to investigate how, why, to what extent, and under what conditions teacher candidates learn to enact this kind of practice, and the connection to student learning outcomes. The following recommendations for future research include endeavors to improve the validity, utility, and feasibility of the current Teaching Equity Enactment Scenario scale. I also discuss the potential contributions of using the Teaching Equity Enactment Scenario scale in research studies that seek to understand the extent to which

teacher candidates' practice evolve as a result of elements within and arrangement of their preparation program.

The first recommendation is to improve the current scale by eliminating the remaining redundancy and stretching up the upper boundary of the scale. For example, referring back to Figure 4.8 (page 138), further item revisions can include removing one scenario each from the two pairs of scenarios (i.e., Scenarios F236M & F346M; and, Scenarios F156M & F456M), since each pair of scenarios shared the same location on the scale. One of the four low-level scenarios that were closely located with each other (i.e., Scenarios F346L, F126L, F156L, and F236L) can also be eliminated based on evidence and thoughtful considerations. One of the two high-level scenarios (i.e. Scenarios F156H and F456H) that shared the same scale location can be revised to be harder with the purpose of stretching up the upper boundary of the scale. The aim of the revisions is to produce a result in a ladder-like uniform, continuous scenario structure defining levels of enactment of teaching practice for equity. The revisions, which involve carefully eliminating three items while ensuring the proper coverage of construct, can serve to make the instrument more feasible and less overwhelming to future participants.

With the revised scenario-scale, the second recommendation is a validation analysis. The purpose of the validation analysis is to examine the extent to which the responses of the Teaching Equity Enactment Scenario scale are related to other scales that measure a theoretically similar construct to teaching practice for equity. For example, the Learning to Teach for Social Justice-Beliefs scale (Ludlow et al., 2008) can be one measure used for this validation study. The results of the validation study can provide empirical evidence for concurrent validity, which contributes to the validity of this instrument.

The third recommendation for future research is to revise the self-report type of

instrument to an observational tool. Mentors or collaborating teachers can use the instrument to observe teacher candidates' or colleagues' practice in the classroom. Teacher candidates or teachers themselves can use the self-report version of the instrument to reflect on their own teaching practice. Correlations between scores obtained from both the observational and the self-report instruments can show how reliably teachers' self-report about their practice for equity predict their classroom practice assessed by their colleagues. Teacher candidates and their mentors or teachers and their colleagues can use the scenarios as a platform to discuss, elaborate, and reflect on their pedagogy, teaching philosophy, instructional purposes, challenges encountered, confusions remained, and suggestions to deal with certain challenges in their classroom and school contexts. The scenario instrument can be a springboard for productive discussion among teachers with the intention of improving teaching practice.

Once the current instrument is further revised and validated, the scenario instrument, including both the self-report and the observational tool, can be useful to diagnose a teacher's level of enacting equity-centered teaching practice and provide a meaningful interpretation. Moreover, this instrument can be used to measure the extent to which teacher candidates' practice change at the beginning of a preparation program, at the end of the program, and a few years into teaching. If the instrument is sensitive enough to capture the growth at different time points, the information offers empirical evidence about one outcome measure of a given preparation program.

### **Implications and Conclusion**

The successful development of a scenario-style scale that measures the complex construct of enactment of practice for equity in a reliable, valid, and authentic manner has implications for research on preparing and supporting teachers to enact equity-centered practice in a democratic

society. The instrument itself offers implications for the field of instrument development and measurement.

This study takes the stance that in a democratic society, the goal of education is to ensure that all citizens are equipped to perceive, reason, and judge information and knowledge, respect diversity, and participate fully in deliberate dialogue and debate for the public good (Guttman, 1987/1999; Sleeter, 2013). The democratic goal of education has become even more critical and urgent given recent political events in the U.S. and in the global context. Given the persistent educational inequalities in opportunities, resources to learning, and learning outcomes, which are a result of and perpetuated by inequitable social, economic, and school systems, many scholars have argued that it is not enough to close the academic achievement gap as measured by standardized test scores. Quality teaching must prepare students to be critical and responsible citizens with the capacity to participate fully in a democratic society (Banks, Cochran-Smith, Moll, Richert, Zeichner, & LePage, 2005). Among all relevant stakeholders who are involved in and responsible for the policy and practice of education, teacher preparation programs hold an important role and responsibility to fulfill the potential of democratic education.

Preparing and supporting teachers to enact teaching practice that responds to diversity, challenges educational inequities, and promotes social justice is a pressing yet daunting and complex task. More research on how to better prepare and support teacher candidates' learning in and beyond preparation programs is needed. Additionally, we need to know how teaching practice for equity influences students' academic, social, emotional, critical and civic learning. One piece of empirical evidence that is needed in conducting studies mentioned earlier is a measure that captures the extent to which teacher candidates or teachers enact teaching practice for equity. The success of developing the scenario-style scale can be used to provide the kind of

evidence needed in research on preparing and supporting teachers to enact teaching with a commitment to equity and social justice. Also, since this instrument can be used to measure change on teaching practice over time, the longitudinal design of studies will help researchers to understand the journey of learning to teach starting from the preparation program through the initial years of teaching. While this instrument can provide the kind of evidence to measure outcomes of teacher preparation, it cannot be used alone as a sole indicator of a preparation program's outcome. Mixed methods and multiple indicators must be used to provide triangulated and valid interpretations of outcomes of teacher preparation programs.

This instrument development study also offers several implications pertinent to the field of measurement. To date, very few instrument-development studies construct scenario-like scales by applying the novel approach of combining Rasch measurement theory and facet theory (Ludlow et al., 2014; Ludlow et al., in press). The promising results of this study again confirm the findings of the previous studies by applying the similar approach of combining Rasch measurement theory and Guttman's facet theory to develop a scenario-scale. Specifically, using facet theory and the sentence mapping techniques to develop items appears to be an efficient and systematic way to ensure content validity of an instrument that measures a complex construct. Moreover, the premises of facet theory and the sentence mapping techniques appropriately reflects the principles of Rasch measurement theory, which considers a construct as a unidimensional, hierarchical continuum. The promising and encouraging results of this study provide the empirical evidence and serve as another example for measurement scholars who are interested in developing a measure that enables meaningful interpretation, objective comparisons, and useful predictions.

Another contribution of this study to the field of measurement and resulting implication is the extensive effort undertaken in item development as detailed in Chapter Three. The promising and encouraging results of the instrument would not exist without the systematic and rigorous processes of construct clarification and conscious decision-making through aligning with the theoretical framework. This suggests that it is worthwhile and productive to obtain a solid understanding of the construct before marching into item development. This dissertation documented the detailed procedure of item development, including critical decisions made at different stages of instrument development, which can provide a useful example of the extensive efforts and steps required to develop a reliable, valid, and useful scale. While the procedures provide a useful example for instrument development, the steps and decisions are not fixed and must be tailor-made according to the construct of interest, the theoretical framework, the purpose of the measure, and the study's context.

This study demonstrated that the combination of Rasch measurement and Guttman's facet theory is successful and productive when developing a scenario-style scale to capture a complex construct – teachers' self-reports about their enactment of practice for equity - in a reliable, valid, and authentic manner. The results and conclusions of this study only mark the need for more meaningful research to understand how programs can better prepare and support teachers to teach with a commitment to equity and social justice in diverse classrooms. This line of research studies not only contribute to the field of research on teaching, but advocate for equity-center social justice teaching, and learning to thrive in a democratic society in a changing and uncertain world.



## References

- Achinstein, B., & Ogawa, R. T. (2012). New teachers of color and culturally responsive teaching in an era of educational accountability: Caught in a double bind. *Journal of Educational Change, 13*(1), 1-39.
- Anderson, D. E., & Maes, W. R. (1986). *Teacher Role Survey* [Database Record]. Retrieved from the Mental Measurements Yearbook.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*, 561-573.
- Andrich, D. (1989). Distinctions between assumptions and requirements in measurement in the social sciences. In J. A. Keats, R. Taft, R. A. Heath, S. H. Lovibond (Eds.), *Mathematical and Theoretical Systems* (pp. 7-16). Elsevier Science Publishers: North Holland.
- Aitken, G., & Sinnema, C. (2008). *Effective pedagogy in social sciences/tikanga a iwi*. Wellington, NZ: Ministry of Education.
- Alton-Lee, A. (2003). *Quality Teaching for diverse students in schooling: Best evidence synthesis*. Wellington, NZ: Ministry of Education.
- Anthony, G., & Walshaw, M. (2007). *Effective pedagogy in mathematics/pangarau*. Wellington, NZ: Ministry of Education.
- Baluch, S., Greig, T., Ponterotto, J., & Rivera, L. (1998). Development and initial score validation of the teacher multicultural attitude survey. *Educational and Psychological Measurement, 58*(6), 1002-1015.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191-215.
- Banks, J., Cochran-Smith, M., Moll, L.C., Richert, A., Zeichner, K.M., & LePage, P. (2005). Teaching diverse learners. In L. Darling-Hammond and J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 232-274). San Francisco, CA: Jossey-Bass.
- Bilsky, W., & Schwartz, S. H. (2008). Measuring motivations: Integrating content and method. *Personal and Individual Differences, 44*, 1738-1751.
- Bishop, R., Berryman, M., Cavanagh, T., & Teddy, L. (2009). Te Kotahitanga: Addressing educational disparities facing Maori students in New Zealand. *Teaching and Teacher Education, 25*, 734-742.
- Blanchard, B. E. (1967). *Illinois Ratings of Teacher Effectiveness Manual. Grades 9-12*. [Database Record]. Retrieved from the Mental Measurements Yearbook.

- Borg, I., & Shye, S. (1995). *Facet Theory*. Thousand Oaks, CA: Sage.
- Boston College Fact Book 2014-15. Retrieved from [http://www.bc.edu/content/dam/files/publications/factbook/pdf/14\\_15/14-15\\_fact\\_book.pdf](http://www.bc.edu/content/dam/files/publications/factbook/pdf/14_15/14-15_fact_book.pdf)
- Brown, J. J. (2012). *The teacher attitudes toward homeless students scale: Development and validation* (Doctoral dissertation). Retrieved from ProQuest. (UMI 3503233)
- Buehler, J., Gere, A. R., Dallavis, C., & Haviland, V. S. (2009). Normalizing the fraughtness. *Journal of Teacher Education*, 60(4), 408-418.
- Center for Advanced Study of Teaching and Learning. (n.d.). Classroom Assessment Scoring System. Retrieved from <http://curry.virginia.edu/research/centers/castl/class>
- Cochran-Smith, M. (1999). Learning to teach for social justice. In G. Griffin (Ed.), *The education of teachers: Ninety-eighth yearbook of the national society for the study of education* (pp. 114–144). Chicago, IL: University of Chicago Press.
- Cochran-Smith, M. (2010). Towards a theory of teacher education for social justice. In M. Fullan, A. Heagreaves, D. Hopkins, & A. Lieberman (Eds.), *The international handbook of educational Change* (2<sup>nd</sup> ed., pp. 445–467). New York: Springer.
- Cochran-Smith, M., Ludlow, L., Ell, F., O’Leary, M., & Enterline, S. (2012). Learning to teaching for social justice as a cross cultural concept: Findings from three countries. *European Journal of Educational Research*, 1(2), 171-198.
- Cochran-Smith, M., Piazza, P., & Power, C. (2013). The politics of accountability: Assessing teacher education in the United States. *The Educational Forum*, 77(1). 6-27.
- Cochran-Smith, M., Ell, F., Ludlow, L., Grudnoff, L., & Aiken, G. (2014a). The challenge and promise of complexity theory for teacher education research. *Teacher College Record*, 116, 1-38.
- Cochran-Smith, M., Ell, F., Grudnoff, L., Ludlow, L., Haigh, M., & Hill, M. (2014b). When complexity theory meets critical realism: A platform for research on initial teacher education. *Teacher Education Quarterly*, 41(1), 105-122
- Cochran-Smith, M., Fiona, E., 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.
- Coster, W., Haley, S. M., Ludlow, L. H., Andres, P. L., & Ni, P. S. (2004). Development of an applied cognition scale to measure rehabilitation outcomes. *Archives of Physical Medicine and Rehabilitation*, 85, 2030-2035.

- Coster, W., Ludlow, L. H., & Mancini, M. C. (1999). Using IRT variable maps to enrich understanding of rehabilitation data. *Journal of Outcome Measurement*, 3, 123-133.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.
- Crocker, L., & Algina, J. (1986). *Introduction to classical & modern test theory*. Mason, OH: Cengage Learning.
- Cronshaw, S. F., & Jethmalani, s. (2005). The structure of workplace adaptive skill in a career inexperienced group. *Journal of Vocational Behavior*, 66, 45-65.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24(4) 349-354. doi: 10.1037/h0047358.
- Cuban, L. (2013). *Inside the black box of classroom practice: Change without reform in American education*. Cambridge, MA: Harvard Education Press.
- Dalton, S. S. (2007). *Five standards for effective teaching: How to succeed with all learners*. San Francisco, CA: Wiley.
- D'Andrea, M., Daniels, J., & Noonan, M. (2003). New developments in the assessment of multicultural competence: The multicultural awareness-knowledge-skills survey-teachers form. In D. B. Pope-Davis, Ed., *Multicultural competencies in counseling and psychology* (pp.154-167). Thousand Oaks, CA: Sage.
- Danielson, C. (2014, August 25). The framework for teaching: Evaluation instrument (2013 Edition). Retrieved from: <https://www.danielsongroup.org/download/?download=448>
- Dover, A. G. (2009). Teaching for social justice and K-12 student outcomes: A conceptual framework and research review. *Equity and Excellence in Education*, 42(4), 506-524.
- Education Testing Center. (n.d.). The *Praxis* Performance Assessment for Teachers from ETS. Retrieved from: <https://www.ets.org/ppa/test-takers/teachers/about>
- 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. *New Educator*, 4(4), 267-290.
- Evans, G. W., Brauchle, G., Haq, A., Stecker, R., Wong, K., & Shapiro, E. (2007). Young children's environmental attitudes and behaviors. *Environment and Behavior*, 39(5), 635-658.

- Fisher, Y. (2013). Exploration of values: Israeli teachers' professional ethics. *Social Psychology Education, 16*, 297-315.
- Fisher, Y. (2014). The timeline of self-efficacy: changes during the professional life cycle of school principals. *Journal of Educational Administration, 52*(1), 58-83.
- Fitchett, P. G., Starker, T. V., & Salyers, B. (2012). Examining culturally responsive teaching self-efficacy in a pre-service social studies education course. *Urban Education, 47*(3), 585-611.
- Freeman, M. (2005). Constant comparative method. In S. Mathison (Ed.), *Encyclopedia of Evaluation*, (pp. 81-82). Thousand Oaks, CA: Sage.
- Freire, P. (1970/2002). *Pedagogy of the oppressed*. New York: Continuum.
- Friedman, I. A. (2004). Organizational expectations of the novice teacher. *Social Psychology of Education, 7*, 435-461.
- Friedman, I. A. (2006). The bi-polar professional self of aspiring teachers: Mission and power. *Teaching and Teacher Education, 22*, 722-739.
- Frye, B., Button, L., Kelly, C., & Button, G. (2010). Preservice teachers' self-perceptions and attitudes toward culturally responsive teaching. *Journal of Praxis in Multicultural Education, 5*(1), 6-22.
- Fulmer, G. W. (2015). Validating proposed learning progressions on force and motion using the force concept inventory: Findings from Singapore secondary schools. *International Journal of Science and Mathematics Education, 13*, 1235-1254.
- Funk, J., Fox, C., Chan, M., & Curtiss, K. (2008). The development of the children's empathic attitudes questionnaire using classical and rasch analyses. *Journal of Applied Developmental Psychology, 29*(3), 187-196.
- Gable, R. K., Ludlow, L. H., McCoach, D. B., & Kite, S. L. (2011). Validation of the survey of knowledge of Internet risk and Internet behavior. *Educational and Psychological Measurement, 71*(1), 217-230.
- Gay, G. (2000). *Culturally responsive teaching: Theory, research & practice*. New York: Teacher College Press.
- Grant, C. A., & Agosto, V. (2008). Teacher capacity and social justice in teacher education. In M. Cochran-Smith, S. Feiman-Nemser, D. J. McIntyre., & K. E. Demers (Eds.), *Handbook of research on teacher education: Enduring questions in changing contexts* (3rd ed., pp. 175–200). New York: Routledge.

- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 11*(3), 255–274.
- Grudnoff, L., & Ell, F., (2013). The politics of responsibility: Teacher education and persistent underachievement. *The Educational Forum, 77*(1), 73-86.
- Grudnoff, L., Haigh, M., Hill, M., Cochran-Smith, M., Ell, F., & Ludlow, L., (2017). *Teaching for equity: Insights from international evidence with implications for a teacher education curriculum. The Curriculum Journal, 1-22.*
- Guttman, A. (1987/1999). *Democratic Education: With a New Preface and Epilogue.* Princeton, NJ: Princeton University Press.
- Guttman, R., & Greenbaum, C. W. (1998). Facet Theory: Its development and current status. *European Psychologist, 3*(1), 13-36.
- Guyton, E. M., & Wesche, M. V. (2005). The multicultural efficacy scale: Development, item selection, and reliability. *Multicultural Perspectives, 7*(4), 21-29.
- Hambleton, R., & Jones, R. (1993). An NCME instructional module on comparison of classical test theory and item response theory and their application to test development. *Educational Measurement: Issues and Practice, 38-47.*
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7<sup>th</sup> ed.). Upper Saddle River, NJ: Pearson.
- Henry, G. B. (1986). Cultural diversity awareness inventory. Virginia: Hampton University Mainstreaming Outreach Project. (ERIC Document Reproduction Service No. ED282657)
- Hollins, E. R., & Torres Guzman, M. (2005). Research on preparing teachers for diverse population. In M. Cochran-Smith & K. Zeichner (Eds.), *Studying teacher education: The report of the AERA Panel on Research and Teacher Education* (pp. 477-544). Mahwah, NJ: Lawrence Erlbaum.
- Howe, K. (1997). *Understanding equal educational opportunity: Social justice, democracy, and schooling.* New York: Teachers College Press.
- James, M., & Pollard, A. (2011). TLRP’s ten principles for effective pedagogy: Rationale, development, evidence, argument and impact. *Research Papers in Education, 26*(3), 275-328.

- Kukic, S. J., Fister, S. L., Link, D. P., & Freston, J. L. (1989). *The Scales for Effective Teaching* [Database Record]. Retrieved from the Mental Measurements Yearbook.
- Ladson-Billings, G. J. (1995a). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, 32(3), 465-491.
- Ladson-Billings, G. J. (1995b). But that's just good teaching! The case for culturally relevant pedagogy. *Theory Into Practice*, 34(3), 159-165.
- Ladson-Billings, G. J. (1999). Preparing teachers for diverse student populations: A critical race theory perspective. *Review of Research in Education*, 24, 211-247.
- Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in U.S. schools. *Educational Researcher*, 35(7), 3-12.
- Linacre, J. (1994). *Many-facet Rasch measurement*. Chicago: MESA Press.
- Linacre, J.M. (1998). Thurstone thresholds and the Rasch model. *Rasch Measurement Transactions*, 12(2), 634-5.
- Linacre, J. M. (2015). WINSTEPS (Version 3.91.2) [Computer program]. Beaverton, OR: Winsteps.com. Retrieved January 1, 2016. Available from <http://www.winsteps.com/>
- Linacre, J. M. (2016). Winsteps® Rasch measurement computer program User's Guide. Beaverton, Oregon: Winsteps.com
- Lord, F. M., & Novick, M.R. (1968). *Statistical Test Theories of Mental Test Scores*. Addison-Wesley.
- Lord, F. M. (1980). Classical Test Theory—Summary and Perspectives (Ch. 1) in *Applications of Item Response Theory to Practical Testing Problems*. LEA.
- Ludlow, L. H. (1986). Graphic analysis of Item Response Theory residuals. *Applied Psychological Measurement*, 10(3), 217-229.
- 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., & Mahalik, J. R. (2001). Congruence between a theoretical continuum of masculinity and the Rasch model: Examining the conformity to masculine norms inventory. *Journal of Applied Measurement*, 2, 205-221.
- Ludlow, L. H., Enterline, S. E., & Cochran-Smith, M. (2008). Learning to teach for social justice-Beliefs Scale: An application of Rasch Measurement Principles. *Measurement and Evaluation in Counseling and Development*, 40(4), 194-214.

- Ludlow, L. H., Rollison, J., Cronin, J., & Wallingford, T. (2012). Development of the teaching economic literacy: Confidence and anxiety (TELCA) instrument. *International Journal of Educational and Psychological Assessment*, 9(2), 82-103.
- 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. (In press). Enhancement and validation of the productive engagement portfolio-scenario (PEP-S8) scales. *Measurement and Evaluation in Counseling and Development*.
- Marshall, P. L. (1996). Multicultural teaching concerns: New dimensions in the area of teacher concerns research? *The Journal of Educational Research*, 89(6), 371-379.
- Maslovaty, N., Marshall, A. E., & Alkin, M. C. (2001). Teachers' perceptions structured through facet theory: Smallest space analysis versus factor analysis. *Educational and Psychological Measurement*, 61(1), 71-84.
- McCarney, S. B., & Cummins, K. K. (1986). *The Teacher Evaluation Scale* [Database Record]. Retrieved from the Mental Measurements Yearbook.
- MET Project. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET project's three-year study*. Bill and Melinda Gates Foundation. Retrieved from <http://files.eric.ed.gov/fulltext/ED540958.pdf>
- Mitchell, K. (2013). Race, difference, meritocracy, and English: Majoritarian stories in the education of secondary multilingual learners. *Race Ethnicity and Education*, 16(3), 339-364.
- National Center for Education Statistics (2013). *The Nation's Report Card: A First Look: 2013 Mathematics and Reading (NCES 2014-451)*. Institute of Education Sciences, U.S. Department of Education, Washington, D.C.
- Nieto, S. (2000). Placing equity front and center: Some thoughts on transforming teacher education for a new century. *Journal of Teacher Education*, 51(3), 180-87.
- No Child Left Behind Act of 2001. (2002). Pub. L. No. 107–110, 115 Stat. 1425.
- Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw Hill.
- O'Leary, M., Lysaght, Z., & Ludlow, L. H. (2013). A measurement instrument to evaluate teachers' assessment for learning classroom practices. *The International Journal of Education and Psychological Assessment*, 14(2), 40-60.

- Piburn, M., & Sawada, D. (2000). Reformed Teaching Observation Protocol (RTOP): Reference Manual. Arizona Collaborative for Excellence in the Preparation of Teachers. ACEPT Technical Report No. IN00-3
- Picower, B. (2009). The unexamined Whiteness of teaching: How White teachers maintain and enact dominant racial ideologies. *Race Ethnicity and Education, 12*(2), 197-215.
- Randall, J., & Engelhard, G. (2010). Using Guttman's mapping sentences and Many Facet Rasch Measurement theory to develop an instrument that examines the grading philosophies of teachers. *Journal of Applied Measurement, 11*(2), 122-141.
- Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B. D. Wright. Chicago, IL: The University of Chicago Press.
- Rasch G. (1966). An individualistic approach to item analysis. In P. F. Lazarsfeld & N. W. Henry (Eds.), *Readings in mathematical social science*, (pp. 84-108). Chicago, IL: Science Research Associates.
- Rettig, K. D., & Leichtentritt, R. D. (1999). A general theory for perceptual indicators of family life quality. *Social Indicator Research, 47*, 307-342.
- Rice, J. K. (2015). Investing in Equal Opportunity: What Would it Take to Build the Balance Wheel? Boulder, CO: National Education Policy Center. Retrieved May 21<sup>st</sup> 2016 from <http://nepc.colorado.edu/publication/balance-wheel>
- Ritter, J. M., Boone, W. J., & Rubba, P. A. (2001). Development of an instrument to assess prospective elementary teacher self-efficacy beliefs about equitable science teaching and learning. *Journal of Science Teacher Education, 12*(3), 175-198.
- Rucker, B. A. (2004). Community awareness project: A tool for preparing culturally sensitive teachers. *Mid-Western Educational Researcher, 17*(2), 12-20.
- Sahlberg, P. (2011). *Finnish lessons*. New York, NY: Teachers College Press.
- Sahlberg, P. (2014). *Finnish lessons 2.0: What can the world learn from educational change in Finland?* New York, NY: Teachers College Press.
- Shireen Desouza, J. M., Boone, W. J., & Yilmaz, O. (2004). A study of science teaching self-efficacy and outcome expectancy beliefs of teachers in India. *Science Education, 88*(6), 837-854.
- Siwatu, K. O. (2007). Preservice teachers' culturally responsive teaching self-efficacy and outcome expectancy beliefs. *Teaching and Teacher Education, 23*, 1086-1101.



- Siwatu, K. O. (2009). Student teachers' self-efficacy beliefs regarding culturally responsive teaching and their professed classroom practices. *Teacher Education and Practice*, 22(3), 323-333.
- Siwatu, K. O. (2011). Preservice Teachers' culturally responsive teaching self-efficacy-forming experiences: A mixed methods study. *Journal of Educational Research*, 104(5), 360-369. doi: 10.1080/00220671.2010.487081
- Sleeter, C. E., & Grant, C. (1987). An analysis of multicultural research in the United States. *Harvard Educational Review*, 57(4), 421–445.
- Sleeter, C. E. (2001). Preparing teachers for culturally diverse schools: Research and the overwhelming presence of Whiteness. *Journal of Teacher Education*, 52(2), 94-106.
- Sleeter, C. E. (2013). *Power, teaching, and teacher education: Confronting injustice with critical research and action*. New York, NY: Peter Lang Publishing.
- Smith C. E., & Andrich D. (2005, April). *Why Rasch Measurement or Why don't (some) psychometricians get along?* Paper presented at the annual meeting of the American Educational Research Association, Montreal, CA.
- Smith, R.M., Linacre, J.M., and Smith, Jr., E.V. (2003). Guidelines for Manuscripts. *Journal of Applied Measurement*, 4, 198-204.
- Soares, L. M., & Soares, A. T. (1999). *The Teacher Performance Assessment* [Database Record]. Retrieved from the Mental Measurements Yearbook.
- Strahan, R., and Gerbasi, K.C. (1972), Short, Homogeneous Versions of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, 28, 191–193.
- United Nations Educational, Scientific, and Cultural Organization (2015). Education For All Global Monitoring Report.
- U.S. Department of Education (2016). Preparing and credentialing the nation's teachers: The secretary's tenth report on teacher quality.
- Ukpokodu, O. N. (2011). How do I teach mathematics in a culturally responsive way? Identifying empowering teaching practices. *Multicultural Education*, 19(3), 47-56.
- Villegas, A.M., & Lucas, T. (2002). *Educating culturally responsive teachers: A coherent approach*. Albany, NY: State University of New York Press.
- Vuilleumier, C., Klein, K, & Ludlow, L. H. (2015, April). *Measuring Social Justice Teaching Beliefs in Pre-service Educators Using Rasch-Based Scenario Scales*. Paper session presented at the meeting of the New England Educational Research Organization, Portsmouth NH.

- Walker-Dalhouse, D., & Dalhouse, D. A. (2006). Investigating White preservice teachers' beliefs about teaching in culturally diverse classrooms. *Negro Educational Review*, 57(1-2), 69-84.
- Waugh, R. F. (2003). Measuring attitudes and behaviors to studying and learning for university students: A rasch measurement model analysis. *Journal of Applied Measurement*, 4(2), 164-180.
- Welner, K., & Farley, A. (2010). *Confronting systemic inequity in education: High impact strategies for philanthropy*. National Committee for Responsive Philanthropy (NCRP). Retrieved from [https://www.ncrp.org/files/publications/Confronting\\_Systemic\\_Inequities\\_in\\_Education\\_lowres.pdf](https://www.ncrp.org/files/publications/Confronting_Systemic_Inequities_in_Education_lowres.pdf)
- Wilson, M. (2004). *Constructing measures: An item response modeling approach*. New York, NY: Psychology Press.
- Wilson, M., Allen, D. D., & Li, J. C. (2006). Improving measurement in health education and health behavior research using item response modeling: Comparison with the classical test theory approach. *Health Education Research*, 21, 19-32.
- Wolfe, E. W., & Smith, E. V. (2007a). Instrument development tools and activities for measure validation using Rasch models: Part I—Instrument development tools. *Journal of Applied Measurement*, 8(1), 97-123.
- Wolfe, E. W., & Smith, E. V. (2007b). Instrument development tools and activities for measure validation using Rasch models: Part II—Validation activities. *Journal of Applied Measurement*, 8(2), 204-234.
- Wright B. D. (1967). Sample-free test calibration and person measurement. In *Proceedings of the 1967 Invitational Conference on Testing Problems*. Princeton: Educational Testing Service, 85-101.
- Wright B. D. (1980). “Foreward”, and “Afterward”. In Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. University of Chicago Press.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA Press.
- Writer, J. H. (2008). Unmasking, exposing, and confronting: Critical race theory, tribal critical race theory and multicultural education. *International Journal of Multicultural Education*, 10(2), 1-15.
- Ysseldyke, J. E., Samuels, S. J., & Christenson, S. L. (1988). Teacher Evaluation Rating Scales [Database Record]. Retrieved from the Mental Measurements Yearbook.

Zeichner, K. (1993). Connecting genuine teacher development to the struggle for social justice. *Journal of Education for Teaching*, 19(1), 5–21.

Zeichner, K. (2005). A research agenda for teacher education. In M. Cochran-Smith, & K. M. Zeichner (Eds.), *Studying teacher education: The report of the AERA panel on research and teacher education* (pp. 737–759). Mahwah, NJ: Erlbaum and Associates.

## Appendix A: Construct Mapping Sheet

Facets	I. Selecting worth while content and designing and implementing learning opportunities aligned to valued outcomes	II. Connecting to students as learners and their lives and experiences	III. Creating learning-focused, respectful and supportive learning environments	IV. Using evidence to scaffold learning and improve teaching	V. Taking agency/responsibility for further professional engagement and learning	VI. Challenging and Recognizing inequities
<b>Frameworks: Danielson's Framework for Teaching</b>						
			2a. "Creating an environment full of respect and rapport, which includes teachers interactions with students through words and action as well as students' interactions with each other. Indicators include..."			
			2b. "Establishing culture for learning which are demonstrated through conveying the important educational value of content and of learning, setting high expectation for learning and achievement, and students showing pride in their work. Indicators include..."			
			2c. "Managing classroom			

		<p>procedures which are indicated through teachers' ability to manage instructional groups efficiently and effectively (purposeful and cooperative group atmosphere), smooth/efficient transition between activities, good management/preparation of materials and supplies, and maximum of instructional time than classroom routine (e.g., taking attendance). Indicators include..."</p>			
		<p>2d. "Managing student behaviors which is demonstrated through setting clear expectation of students' conduct, monitoring students' behaviors (as if teachers have eyes at the back of their head), and responding to students' misbehavior early in a manner that</p>			

			respecting students' dignity. Indicators include..."			
			2e. "Organizing physical space, including safety and accessibility of classroom, arranging furniture and using physical resource to maximize learning. Indicators include..."			
	3a. "Communicating with students which is characterized by teachers' ability to communicate the learning goal to students clearly, provide clear instructions/directions for in-class activities, explain concepts/strategies to students by using vivid language, imaginative metaphors/analogies, and connecting to students' lives/experiences, use clear/precise/accurate/expressive oral or written language for					

	students' to emulate. Indicators include..."					
	3b. "Using questions and discussion technique which are demonstrated by teachers' ability to pose quality questions/prompts that cause students to think and reflect and deepen their understanding (through low-cognitive and high-cognitive questions), to use discussion technique to challenge students to examine their premise/build a logical argument/critique the arguments of others, to formulate further discussion questions by building on students' response, and to use a range of techniques to encourage all students to contribute to the discussion.					

	Indicators include..."					
	3c. "Engaging students in learning is shown when students are intellectually active in learning important and challenging content but not rote learning. This can be observed based on what students are asked to do (i.e., activities/assignments that involves thinking and explaining their thinking process), purposeful grouping of students, selection of instructional materials and resources that are better suited to engaging students in deep learning, and well-designed lesson structure/pacing that allows time for reflection and closure. Indicators include..."					
				3d. "Using assessment in instruction as the assessment"		



				<p>t for learning and monitoring students' understanding. This can be demonstrated through setting clear assessment criteria, monitoring students learning carefully in a way that is integrated well into lesson, providing timely/constructive/substantive feedback to guide/advance students' learning, creating clear criteria that promote students' self-assessment/monitoring of their own progress. Indicators include...</p>		
--	--	--	--	---	--	--

				<p>3e.  “Demonstrating flexibility and responsiveness refer to a teacher's skill in making adjustments in a lesson to respond to changing conditions . This subdomain include elements like making minor/(at times) major adjustment to a lesson when needed, responding to students and seizing the "teachable moment" during the class, being persistent in seeking alternate approaches to help students learn. Indicators include...”</p>		
<b>Framework: Teaching and Learning Research Program (TLRP)</b>						

<b>TLRP</b>	<p>1. Effective pedagogy equips learners for life in its broadest sense: learning is defined broadly as developing one's intellectual, personal, and social resources that will enable them to participate as active citizens, contribute to economic development, and thrive in a diverse/changing society. <i>The learning outcome must also take seriously issues of equity and social justice for all.</i></p>					
	<p>2. Effective pedagogy engages with valued forms of knowledge: Pedagogy should engage learners with the big ideas, key skills and processes, modes of discourse, ways of thinking and practicing, attitudes and relationships, which are the most valued learning processes and outcomes in particular contexts.</p>					

	<p>There is indeed some knowledge to be learned, but at the same time it is equally important to learn that valued knowledge is produced, contested, and changed in dialogical processes within and between communities of practice.</p>					
		<p>3. Effective pedagogy recognizes the importance of prior experience and learning: Pedagogy should take account of what the learner knows already in order for them, and those who support their learning, to plan their next steps. Prior learning includes personal and cultural experiences of different groups of learners.</p>				
				<p>4. Effective pedagogy requires learning to be scaffolded</p>		

				<p>: Teachers, trainers and all those, including peers, who support the learning of others, should provide activities, cultures and structures of intellectual, social and emotional support to help learners to move forward in their learning.</p>		
				<p>5. Effective pedagogy needs assessment to be congruent with learning: Assessment should be designed and implemented with the goal of achieving maximum validity both in terms of learning outcomes and learning</p>		

				processes.		
			<p>6. Effective pedagogy promotes the active engagement of the learner: This involves acquiring a repertoire of learning strategies and practices, developing positive learning dispositions, and having the will and confidence to become independent agents in their own learning. The promotion of learner independence and autonomy, in this sense, is not just about the effectiveness of learning. It also concerns the realization of rights, formation as a person, manifestation of citizenship and contribution of individuals to history.</p>			
			<p>7. Effective pedagogy fosters both individual and social processes and outcomes: Learners should be</p>			

			encouraged and helped to build relationships and communication with others for learning purposes, in order to assist the mutual construction of knowledge and enhance the achievements of individuals and groups.			
		8. Effective pedagogy recognizes the significance of informal learning: Informal learning, such as learning out of school or away from the workplace, should be recognized as at least as significant as formal learning and should therefore be valued and appropriately utilized in formal processes.				
					9. Effective pedagogy depends on the learning of all those who support the learning of others: The need for lecturers, teachers, trainers and	

					co-workers to learn continuously in order to develop their knowledge and skill, and adapt and develop their roles, especially through practice-based inquiry, should be recognized and supported.	
					10. Effective pedagogy demands consistent policy frameworks with support for learning as their primary focus: Organizational and system level policies need to recognize the fundamental importance of continual learning - for individual, team, organizational and system success - and be designed to create effective learning environments for all	



					learners.	
--	--	--	--	--	-----------	--

## Appendix B: Collection of Practice for Six Facets

### Facet One

<b>Facet I. Selecting worth while content and designing and implementing learning opportunities aligned to valued outcomes</b>
<b>MET FFT</b>
<p>3a. Communicating with students which is characterized by teachers' ability to communicate the learning goal to students clearly, provide clear instructions/directions for in-class activities, explain concepts/strategies to students by using vivid language, imaginative metaphors/analogies, and connecting to students' lives/experiences, use clear/precise/accurate/expressive oral or written language for students' to emulate. Indicators include clarity of lesson purpose, clear directions and procedures specific to the lesson activities, absence of content errors and clear explanations of concepts and strategies, and correct/imaginative use of language.</p>
<p>3b. Using questions and discussion technique which are demonstrated by teachers' ability to pose quality questions/prompts that cause students to think and reflect and deepen their understanding (through low-cognitive and high-cognitive questions), to use discussion technique to challenge students to examine their premise/build a logical argument/critique the arguments of others, to formulate further discussion questions by building on students' response, and to use a range of techniques to encourage all students to contribute to the discussion. Indicators include questions of high cognitive challenge, formulated by both students and teacher; questions with multiple correct answers or multiple approaches, even when there is a single correct response; effective use of student responses and ideas; discussion, with the teacher stepping out of the central, mediating role; focus on the reasoning exhibited by students in discussion, both in give-and-take with the teacher and with their classmates; high levels of student participation in discussion.</p>
<p>3c. Engaging students in learning is shown when students are intellectually active in learning important and challenging content but not rote learning. This can be observed based on what students are asked to do (i.e., activities/assignments that involves thinking and explaining their thinking process), purposeful grouping of students, selection of instructional materials and resources that are better suited to engaging students in deep learning, and well-designed lesson structure/pacing that allows time for reflection and closure. Indicators include student enthusiasm, interest, thinking, and problem solving; learning tasks that require high-level student thinking and invite students to explain their thinking; students highly motivated to work on all tasks and persistent even when the tasks are challenging; students actively "working", rather than watching while their teacher "works"; suitable pacing of the lesson: neither dragger out nor rushed, with time for closure and student reflection.</p>
<b>BES – Teaching Diverse Students</b>

1. Quality teaching is focused on raising student achievement (including social outcomes) and facilitate high standards of student outcomes for diverse learners. Quality teaching should facilitate the learning of diverse students and raises achievement for all learners. Teachers need to hold high expectations to all learners and recognize that the idea of learning styles approach has led to inappropriate stereotyping of minority students by teachers (e.g. minority students as kinaesthetic learners). Research suggests that the narrower, procedural hands-on approach was less effective in supporting student learning than a conceptual and child-centered approach that highlights children's thinking through interaction with equipment and emphasizes the interrelationships between mathematical ideas. That is, inappropriate assumptions about ethnically-based learner needs can maintain the least effective learning opportunities for these students. To achieve the goal of raising achievement for all students, teachers establish and follow through on appropriate expectations for learning outcomes (e.g., through properly designed diagnostic assessment to facilitate learning) and the pace at which learning should proceed. Evidence suggests that while teacher expectations for high standards are necessary, expectations alone are insufficient to facilitate achievement when not supported by quality teaching. High expectations need to be integrated into quality teaching practices and demonstrates clear links between pedagogical practices and achievement outcomes. Also, attention should be paid to the development and precision in the language and practice of diagnostic assessment. Diagnostic or formative assessment can play an integral role in assisting the teacher to raise student achievement as long as the assessment practices are integrally embedded within, and facilitative of quality teaching, rather than an alternative focus.

5. Opportunity to learn is effective and sufficient: Quality teaching provides sufficient and effective opportunity to learn, and includes and optimizes the effective use of non-linguistic representations by teacher and students. Teachers facilitate students' learning rather than emphasize the compliance of behavior. Curriculum has coherence and interconnectedness to students' real life experiences and address diversity appropriately and effectively. Students are given space to resolve cognitive conflicts in classrooms and have sufficient and appropriate opportunities for practice and application.

6. Multiple task contexts support learning cycles: Task cycles match developmental learning cycles of students; Task cycles enable students to engage in and complete learning processes so that what is learned is remembered; Optimal use is made of complementary combinations of teacher-directed groupings, co-operative groups, structured peer interaction and individual work (including homework) to facilitate learning cycles.

7. Curriculum goals, resources including ICT usage, task design, teaching and school practices are effectively aligned: Curricular alignment should optimize rather than inhibits critical thinking. Curriculum alignment include: the use of resources, teaching materials and ICT is aligned with curriculum goals to optimize student motivation and accomplish instructional purposes and goals; pedagogical strategies are evaluated in relation to curricular goals; ICT usage is integrated into pedagogical practice across the curriculum. Quality teaching is optimized when there is whole school alignment around evidence-based practices. Whole school alignment supports effective inclusion

of diverse students within the school community, supports teachers in maintaining, school-home partnerships focused on learning, minimizes disruptions to quality teaching and sustains continuous improvement, optimizes opportunity, enables a common language, teacher collaboration and reflection and other synergies around improving teaching.

**BES – Social Science**

2.1 Aligning experiences to important outcomes through first identifying prior knowledge: Using research-based trajectories of students conceptual and skill development as points of reference rather than a fixed knowledge to understand the possible nature of students' prior knowledge; employing appropriate and a wide range of techniques/strategies rather than a single/narrowly-diagnostic strategy for accessing to certain type of prior knowledge; recognizing teachers' own prior knowledge (e.g., sufficient subject content knowledge, biased/discriminatory prior knowledge - teachers' prior knowledge can influence their sense of agency when it comes to disrupting a dominant discourse and their ability to make judgment on selecting content and resources) and its impact on the extent to which they can support student learning. Identification of students' prior knowledge helps teachers set the direction for learning by distinguishing "new" learning from what is already known, alerts teachers to the transfer of existing understandings that may inhibit new learning, alerts teachers to student misunderstandings that may inhibit new learning.

2.2 Aligning experiences through purposefully aligning activities and resources (i.e., spoken, written, visual, and experiential) to intended outcomes and the alignment must be continuously checked against the intended outcomes: Practices that purposefully align activities to desired outcome include: 1) expected understanding is explicitly included in the learning experiences; 2) expected understanding is implicitly included or partially embedded in the learning experiences; 3) the learning experiences incorporates additional information, explanation, and examples; 4) preparatory activity or discussion, and contextual information describing the focus or purpose of the activity are included. According to studies cited in the BES, these strategies seem to be consistently effective across different learning outcomes - use of strategies, development of conceptual knowledge, attitude and behavior, and social skills. Research findings also suggest that it may be necessary to align not only the instruction and the intended outcome but also the instruction and learner prior knowledge. Resources (e.g., spoken, written, visual, and experiential) should also be aligned to the purpose of a task that supports students in achieving outcomes related to that task. Teachers can also model the intended outcomes which makes alignment transparent to students, make the purpose of tasks clear to students (explain how the task will be assessed and provide resources such as templates, planning/writing guides, and structured overviews that scaffold them through the task) reducing ambiguity and helps them to focus on important learning. Teacher should also align assessment with teaching which helps communicate what is important.

2.3 Aligning experiences through providing multiple and frequent opportunities to revisit concepts and learning processes: Previous research suggests that students need 3 to 5 aligned experiences and each new occasion should not be separated from the previous occasion by more than two days. Teachers should provide multiple learning opportunities to support students' concept development. Research suggests that the

most successful teaching was characterized by: carefully structured activities that allowed for sustained interactions between partners of differing levels of ability; intentional repetition of the same information from different sources (video and magazine); use of empathy questions (such as ‘How would you feel if ...?’) to help students personalize historical events; use of brief discussion (questions asked during the video showing, which appear not to have interrupted the flow of content); and the opportunity for students to demonstrate understanding through oral interviews. Teachers should provide opportunities for repeated practice to support skill development, and reduce coverage to enable students to focus on important ideas and processes.

2.4 Aligning experiences through attending to the learning of individual students: The overwhelming conclusion from previous research was that prior knowledge and learning are unique to each student. Therefore, teachers need to understand that differing levels of engagement with relevant ideas lead to differences in outcomes for students and differences in interests, involvement, and background knowledge also influence students' engagement. The relevance of this to teachers lies in the power they have to connect learning to prior knowledge, to relate content to students' interests, and to promote students' involvement and participation (as well as their engagement with relevant, aligned content) in relevant activities. Teachers support learning by attending to the social and cognitive aspects of the learning process when making teaching decisions. Suggested practices from research include: 1) design tasks that engage students' interest and focus their mind on relevant content (tasks that are both motivating and aligned); 2) make connections with the realities of students' lives; 3) minimize peripheral activities, such as drawing margins, pasting, and coloring headings; 4) constantly monitor pupil involvement and avoid and correct misunderstandings; 5) attend to the social interactions that take place in learning experiences, recognizing how social dominance, status, ability, and knowledge collectively impact on students' learning about their abilities as well as their learning of the content; 6) structure tasks that enable students to think seriously about content for sufficient time. Finally, the sequence in which different activity types occur can also influence what students learn from each activity, especially where the concept was more complex.

### **CREDE**

#### **4. Challenging Activities: Teaching Complex Thinking — Challenge students toward cognitive complexity:**

- 4.1. assures that students — for each instructional topic — see the whole picture as a basis for understanding the parts.
- 4.2. presents challenging standards for student performance.
- 4.3. designs instructional tasks that advance student understanding to more complex levels.
- 4.4. assists students to accomplish more complex understanding by building from their previous success.
- 4.5. gives clear, direct feedback about how student performance compares with the challenging standards.

### **Te Kotahitanga**

2. *Manaakitanga*: Teachers care for the performance of their students

**TLRP**

1. Effective pedagogy equips learners for life in its broadest sense: learning is defined broadly as developing one's intellectual, personal, and social resources that will enable them to participate as active citizens, contribute to economic development, and thrive in a diverse/changing society. The learning outcome must also take seriously issues of equity and social justice for all.

2. Effective pedagogy engages with valued forms of knowledge: Pedagogy should engage learners with the big ideas, key skills and processes, modes of discourse, ways of thinking and practicing, attitudes and relationships, which are the most valued learning processes and outcomes in particular contexts. There is indeed some knowledge to be learned, but at the same time it is equally important to learn that valued knowledge is produced, contested, and changed in dialogical processes within and between communities of practice.

*Facet Two*

<b>Facet II. Connecting to students as learners and their lives and experiences</b>
<b>MET FFT</b>
<b>MET CLASS</b>
Domain 1. Emotional Support: Teacher sensitivity, Regard for students
<b>MET TRIPOD</b>
Care: show concern and commitment
Captive: Inspire curiosity and interest
<b>BES – Teaching Diverse Students</b>
3. Effective links are created between school cultural contexts and other cultural contexts in which students are socialized to facilitate learning: In summary, teachers need to ensure that student experiences of instruction have known relationships to other cultural contexts in which the students are socialised. They need to have the relevance of their learning activities made transparent. Also, the nature of parental or caregiver involvement in their children's education is crucial to improved outcomes. A key research finding is that school-home partnerships are critically dependent upon the agency of educators, their ability to avoid deficit or stereotypical characterisations of parents and caregivers, and their ability to initiate links, respond to, and recognise strengths within the diverse families of their students. Partnerships that align school and home practices and enable parents to actively support their children's in-school learning have shown some of the strongest impacts on student outcomes, especially in literacy and health and physical education. Such partnerships can provide a particularly cost-effective approach to supporting the learning of diverse students. Relevance is made transparent to students. Cultural practices at school are made transparent and taught. Ways of taking meaning from text, discourse, numbers or experience are made explicit. Quality teaching recognises and builds on students' prior experiences and knowledge. New information is linked to student experiences. Student diversity is utilised effectively as a pedagogical resource.
<b>BES – Social Studies</b>
1.1 Making connection to students' life by drawing on relevant content (refer to Gay Geneva's 5 components of Culturally Responsive Teaching Pedagogy): First, teachers should embed students' cultural knowledge and experiences in content that enables cultural continuity. Some teaching practices/strategies include targeting various perspective (i.e., aboriginal and non-aboriginal) when planning learning outcomes, integrating aboriginal learning resources (e.g., literature) into learning program, including resources that provide counter stories to those in mainstream textbook to expose, challenge, and criticize the normalized/privileged discourse, integrating pedagogical strategies documented as effective for teaching and learning of aboriginal student, using a variety of approaches (e.g., portfolio, artifacts, as well as traditional tests) for assessment purpose, and using culture/aboriginal philosophy as the underpinning of curriculum design rather than "add-on." Second, teachers can make students' own lives a point of comparison that supports access to new learning through activities of comparison. Instructional strategies such as similarities-and-differences (involving activities such as comparisons, classifying, metaphor, analogy) are found to be effective. For example, teachers can ask students to compare different

accounts/sources to come to understanding different realities (e.g., compare their own and other's culture). However, a similarities-and-differences strategy needs to be used carefully so that it does not promote binary thinking, resulting in students adopting an 'us-other' or 'them-us' dichotomy. These strategies need to be used in ways that support students to value diversity, rather than entrench racism or promote a 'tourist curriculum'.

1.2 Making connection to students' life by ensuring inclusive content: Teachers should attend to the language that can either make diversity visible or bias student understanding or use resources that can either make diversity visible or bias student understanding. While strategies aimed at encouraging students to make comparisons with their own experiences can support new learning, it is important that teachers attend to the use of language and to content and resource selection to ensure that diversity is not unwittingly excluded. Students in any group are diverse – in terms of gender, ethnicity, culture, disability, age, sexuality, and the like – so if learning is to connect to their lives, resource and content selection needs to reflect this diversity. Where resources fail to make diversity visible, students have fewer opportunities to make connections and create meaning.

4.1 Designing experiences that interest students by meeting diverse motivational needs: Teachers need to recognize that students aren't all motivated by the same type of activities and that students' interest might not match teachers. Therefore, teachers need to find out what motivates their students (e.g., asking students to rank teaching approaches/instructional activities or ask students to log their subjective experiences on different instructional activities every time), allow students to make their own learning choices through student projects (i.e., inquiry-based learning through projects, students as researchers them self generating questions and pursue worthwhile learning for themselves) and stay away from scripted curriculum.

4.2 Designing experiences that interest students by maximizing student interest: Teachers can use first-hand experience of social, cultural, economic, and political situations to make learning real (examples: empowering students to take on real-world problems surrounding their communities and take responsibilities of addressing these issues; creating purposeful and learning-focused out-of-school learning opportunities such as field experiences; having visiting speakers; using drama and simulations, visiting theatre/museums, having hands-on activities ) and to bridge classroom learning with real world life. Teachers can also use narratives that have emotional appeal to engage students or have certain design and selection of resources that have impacts on students' interests (e.g., through illustration/pictures, videos/animation - stop every 4~10 min segment for 3~4 times to clarify content, diagrams, multimedia tools/simulations/games). Teachers should ensure that activities are not only interesting and motivating students, but most importantly learning-focused and aligned to important outcomes.

4.3 Designing experiences that interest students by using a variety of activities: Teachers should use a variety of activity types to help students recall the content



embedded in those experiences.
<b>CREDE</b>
<p><b>2. Language Development: Developing Language Across the Curriculum Develop competence in the language and literacy of instruction across the curriculum:</b></p> <p>2.1. listens to student talk about familiar topics such as home and community.</p> <p>2.2. responds to students’ talk and questions, making ‘in-flight’ changes during conversation that directly relate to students’ comments.</p> <p>2.3. assists written and oral language development through modeling, eliciting, probing, restating, clarifying, questioning, praising, etc., in purposeful conversation and writing.</p> <p>2.4. interacts with students in ways that respect students’ preferences for speaking that may be different from the teacher’s, such as wait-time, eye contact, turn-taking, or spotlighting.</p> <p>2.5. connects student language with literacy and content area knowledge through speaking, listening, reading, and writing activities.</p> <p>2.6. encourages students to use content vocabulary to express their understanding.</p> <p>2.7. provides frequent opportunity for students to interact with each other and the teacher during instructional activities.</p> <p>2.8. encourages students’ use of first and second languages in instructional activities.</p>
<p><b>3. Contextualization Making Meaning: Connecting School to Students’ Lives — Connect teaching and curriculum to students’ experiences and skills of home and community:</b></p> <p>3.1. begins activities with what students already know from home, community, and school.</p> <p>3.2. designs instructional activities that are meaningful to students in terms of local community norms and knowledge.</p> <p>3.3. acquires knowledge of local norms and knowledge by talking to students, parents or family members, community members, and by reading pertinent documents.</p> <p>3.4. assists students to connect and apply their learning to home and community.</p> <p>3.5. plans jointly with students to design community-based learning activities</p> <p>3.6. provides opportunities for parents or families to participate in classroom instructional activities.</p> <p>3.7. varies activities to include students’ preferences, from collective and cooperative to individual and competitive.</p> <p>3.8. varies styles of conversation and participation to include students’ cultural preferences, such as co-narration, call-and-response, and choral, among others.</p>
<b>Te Kotahitagan</b>
1. <i>Manaakitanga</i> : Teachers care for their students as culturally located human beings above all
<b>TLRP</b>
3. Effective pedagogy recognises the importance of prior experience and learning: Pedagogy should take account of what the learner knows already in order for them, and those who support their learning, to plan their next steps. Prior learning includes personal and cultural experiences of different groups of learners.
8. Effective pedagogy recognises the significance of informal learning: Informal

learning, such as learning out of school or away from the workplace, should be recognised as at least as significant as formal learning and should therefore be valued and appropriately utilised in formal processes.

*Facet Three*

<b>Facet III. Creating learning-focused, respectful and supportive learning environments</b>
<b>MET FFT</b>
2a. Creating an environment full of respect and rapport, which includes teachers interactions with students through words and action as well as students' interactions with each other. Indicators include respectful talk, active listening and turn-taking; acknowledgement of students' lives and backgrounds outside of classroom; body language indicative of warmth and caring by both teachers and students; physical proximity; politeness and encouragement; fairness.
2b. Establishing culture for learning which are demonstrated through conveying the important educational value of content and of learning, setting high expectation for learning and achievement, and students showing pride in their work. Indicators include belief in the value of what's been learned; high expectations, supported through both verbal and nonverbal behaviors, for both learning and participation; expectation of high-quality work on the part of students; expectation and recognition of effort and persistence on the part of students; high expectation for expression and work products.
2c. Managing classroom procedures which are indicated through teachers' ability to manage instructional groups efficiently and effectively (purposeful and cooperative group atmosphere), smooth/efficient transition between activities, good management/preparation of materials and supplies, and maximum of instructional time than classroom routine (e.g., taking attendance). Indicators include smooth function of all routines, little or no loss of instructional time, students playing an important role in carrying out the routines, and students knowing what to do and where to move.
2d. Managing student behaviors which is demonstrated through setting clear expectation of students' conduct, monitoring students' behaviors (as if teachers have eyes at the back of their head), and responding to students' misbehavior early in a manner that respecting students' dignity. Indicators include clear standards of conduct, possibly posted and possibly referred to during a lesson; absence of acrimony between teachers and students concerning behaviors; teacher awareness of student conduct; preventive action when needed by the teacher; absence of misbehavior; and reinforcement of positive behaviors.
2e. Organizing physical space, including safety and accessibility of classroom, arranging furniture and using physical resource to maximize learning. Indicators include pleasant/inviting atmosphere, safe environment, accessibility for all students, furniture arrangement suitable for the learning activities, and effective use of physical resources, including computer technology, by both teacher and students.
<b>MET CLASS</b>
Domain 1. Emotional Support: Positive climate, absence of negative climate
Domain 2. Classroom organization: dimension 1~3. Behavior management, productivity, instructional learning formats
<b>MET TRIPOD</b>
Control*: Sustain order, respect, and focus
Confer: Invite ideas and promote discussion

## **BES – Teaching Diverse Students**

2.1 Pedagogical practices create an environment that works as a learning community: The learning community concept has arisen out of the research literature and denotes both a central focus on learning and the interdependence of the social and the academic in optimising learning conditions. The formulation of a "learning community" includes teachers' pedagogical practices that shapes peer culture of supporting each other's learning in the classrooms as well as teachers' role in interacting with students. Teachers can also use instructional organisation and task design to develop learning community. Student friendship patterns and peer status were directly shaped by the teacher's instructional design, task format, task management, the rationale for group membership, and grouping practices. For instance, multi-task environments enable diverse students to variously perform well because the multiple tasks engaged different student strengths at different times. Other practices included: managing resource access, developing students' skills in co-operative group work, developing perseverance in students, and the management of reciprocal problem solving roles with boys and girls of different ethnicities.

2.2 Caring and support is generated through the practices and interactions of teacher(s) and students: learning is supported when structures for caring, opportunities for collaborative learning and appreciation for diversity are established in classrooms. Teacher's care and support is not only manifested through care for students as individual, but also teacher listening, caring about teaching, providing feedback and assistance and so on, as aspects of care.

2.3 Pedagogical practices and class-session to pro-actively value and address diversity: The dual dimensions of care and valuing of diversity are core features of a learning community. Caring practices alone are insufficient to create an environment that supports the learning of diverse students. The diversity of students should be honored as central to quality teaching - it needs to be addressed rather than transcend and effective teaching strategies should build upon the language practices of diverse students within mainstream curriculum rather than as an add-on.

2.4 Academic norms are strong and not subverted by social norms: Mutual respect and trust must be additive to academic or disciplinary norms. It is a balance that teachers manage to maintain when they encourage and support expression of ideas whilst ensuring critical thinking about the ideas. For example, the use of "wait time" enables students opportunities to reflect further and elaborate, and can play in supporting deeper learning, especially for previously low-achieving students. Students also need teacher modelling and explicit teaching that evidence, rather than power, should be used to resolve intellectual conflict in order to genuinely assist the learning process and develop a peer learning culture. The research indicates that, when conceptual disagreements cannot be addressed, learning can be at risk.

2.5 Teaching and tasks are structured to support, and students demonstrate, active learning orientations: Skilled teachers structure and sequence instructional tasks to motivate diverse learners to maintain intellectual engagement with the curriculum content as part of the social interaction of seating or working groups.

2.6 The teacher leads in representing 'us' as everyone in our class community: The teacher has a key role in representing class community to the students, and with the students, in ways that do not exclude by ethnicity, gender, dis/ability, social class

<p>background or sexuality. The concept of learning community moves away from distinctions between the „mainstream“ and the „included disabled“. Rather, difference and diversity become central to our understanding of community.</p>
<p>2.7 Teachers use class sessions to value diversity, and to build community and cohesion: Effective use of teacher-student discussion (e.g., use of wait time or prompts) in whole class contexts enables diversity of student experience to be valued and to be a resource to support student achievement gains. The management of safe participation and avoidance of practices which shame or humiliate students publicly is seen to be important for all students.</p>
<p>2.8 Teachers use instructional organisation and task design to develop learning community: Research suggests that student friendship patterns and peer status were directly shaped by the teacher's instructional design, task format, task management, the rationale for group membership, and grouping practices. Specifically, multi-task environments enable diverse students to variously perform well because the multiple tasks engaged different student strengths at different times. Other practices included: managing resource access, developing students' skills in co-operative group work, developing perseverance in students, and the management of reciprocal problem solving roles with boys and girls of different ethnicities.</p>
<p>2.9 Teachers teach students how to provide help to each other with resource access, dialogue and elaborated explanations: These practices can include fostering peer dialogue that supports cognitive and metacognitive strategies, reciprocal teaching (scaffolding student structured and reciprocal participation in dialogue using questioning, summarising, clarifying and predicting), collaborative reasoning (teachers structure classroom opportunities to optimise discussion amongst peers), fostering a community of learners (teachers use multiple strategies such as jigsaw learning groups to which each child brings a resource needed by others, cross-talk and cross-age tutoring to foster collaboration and the distribution of peer expertise), and computer supported intentional learning environments (support for intentional learning and cognitive strategy use through student use of computer technology).</p>
<p>2.10 Teaching includes training in collaborative group work with individual accountability mechanisms. Students demonstrate effective co-operative and social skills that enable group processes to facilitate learning for all participants: Collaborative group work is defined as 'students working together in a group small enough that everyone can participate on a collective task that is clearly assigned. Research findings suggest that when conceptual learning is the focus, the use of tasks with ill-structured solutions (not simple, structured answers, but real life and complex solutions) can maximise the effectiveness of the co-operative group work.</p>
<p>2.11 Pedagogical practice is appropriately responsive to the interdependence of socio-cultural and cognitive dimensions: Research on the integrating cultural norms into language learning demonstrates the importance of making explicit and developing the socio-cultural norms that support students, not only in strong cultural identity and social development but also in their achievement. A study shows how the teacher's pedagogical skill, particularly in scaffolding student learning, is a key to ensuring a positive and affirming social environment while optimising academic challenge. Another study emphasises the role of social norms in fostering not only cognitive abilities but also the development of “important dispositions, such as students”</p>

willingness to persist in trying to solve difficult problems, and their identities as capable learners.

### **BES – Social Studies**

3.1 Building and sustaining a learning community by establishing productive teacher-student relationships: Characteristics of learning communities include children are listened to, are supported in expressing their views, have their views taken into account, are involved in decision-making processes, and share power and responsibility for decision making. To build a learning community, teachers should build cognitive (for example, responding to ideas in ways that generated new understandings), social (for example, sharing responsibility for learning decisions), emotional (for example, allowing the students to know something about other roles in his life and being open and honest about the emotions involved), and spatial (for example, sitting with the students, at their level) connections with students, which increases students sense of belonging and motivations and involvement of diverse students (diversity is accepted, supported, and valued). Respectful relationships are not only important for the influence they have on motivation, they are also vital for creating a sense of affiliation and belonging – to the class and the school community. Inclusive relationships and value for diversity are shown when teachers treat all students as of equal status, relationships were reciprocal, and children engaged in the full range of roles pertinent to the setting. Teachers can also model learning behaviors in relationships which helps embed a focus on learning. For example, teachers can refocus the children on learning rather than doing by using learning language (e.g., we are learning together, are you switched on for learning) to make learning more visible to the children and to continue bringing learning to the surface.

3.2 Building and sustaining a learning community by promoting dialogue: Research suggests that students learn content when they talk together about that content. Teachers can create the environment and provide infrastructure to allow students' group work, design activities that facilitate shared knowledge building or utilize group learning skills, design complex tasks (open-ended, uncertain, and involve big idea) that require group interactions/collaboration/productive dialogue (i.e., reciprocal rather than sequential interaction) to complete the task. To develop group work norms and foster group functioning, certain desired behaviors must be labeled and discussed with students so students can recognize when it occurs and practice those behaviors. Also, students create and agree on a shared set of ground rules for exploratory talk, to be used when working in groups. These rules are based on the understanding that “high quality speaking and listening is of great value in class; high quality speaking and listening is inclusive and respectful of opinions and ideas; all information is shared; reasons are requested and given and the group seeks to reach agreement. The children’s ownership of the rules helps the groups to implement them. Teachers can provide explicit skills training to students to promote cooperation and dialogue through strategies such as role play or whole-class explicit skill-building intervention (e.g., deliberately listen/speak/ask questions, recognize each other's contribution, give and receive help, deliberately talking about group communication skills, setting ground

rules, reflecting on group work process). More specifically, one skill can be taught each week, supported every day by practice and feedback opportunities to reinforce new learning behaviors when they occur. Teachers could also model interactional skills during cooperative learning to help learners develop dialogue skills, use complex tasks to promote reciprocal rather than sequential interaction, publicly acknowledge multiple abilities/voices among learners, use properly designed whole-class discussion that involve discussion or debates, opportunities to participate in decision making, use statements (i.e., declarative, reflective restatements, or invitations to elaborate statements) instead of questions to stimulate conversations or invite further elaboration.

3.3 Building and sustaining a learning community by sharing power with students: Teaching practices include deliberately delegating authority to students to make decisions about their learning, and – indirectly – teaching them in ways that enable them to be more independent in their learning and less dependent on their teacher (e.g., teach both content and learning strategies that are transferable to other situations; support learners to take more responsibility for their own learning through the use of regular, structured, reflective activities; provide a metacognitive environment that integrate direct teaching of thinking skills and ‘thinking journals’, in which students responded to metacognitive prompts; apply principles of both generative teaching and cooperative teaching approach (see pp. 177).

### **CREDE**

#### **1. Joint Productive Activity:**

- 1.1. designs instructional activities requiring student collaboration to accomplish a joint product.
- 1.2. matches the demands of the joint productive activity to the time available for accomplishing them.
- 1.3. arranges classroom seating to accommodate students’ individual and group needs to communicate and work jointly.
- 1.4. participates with students in joint productive activity.
- 1.5. organizes students in a variety of groupings, such as by friendship, mixed academic ability, language, project, or interests, to promote interaction.
- 1.6. plans with students how to work in groups and move from one activity to another, such as from large group introduction to small group activity, for clean-up, dismissal, and the like.
- 1.7. manages student and teacher access to materials and technology to facilitate joint productive activity.
- 1.8. monitors and supports student collaboration in positive ways."

<p><b>5. Instructional Conversation: Teaching Through Conversation — Engage students through dialogue, especially the Instructional Conversation:</b></p> <p>5.1. arranges the classroom to accommodate conversation between the teacher and a small group of students on a regular and frequent basis.</p> <p>5.2. has a clear academic goal that guides conversation with students.</p> <p>5.3. ensures that student talk occurs at higher rates than teacher talk.</p> <p>5.4. guides conversation to include students’ views, judgments, and rationales using text evidence and other substantive support.</p> <p>5.5. ensures that all students are included in the conversation according to their preferences.</p> <p>5.6. listens carefully to assess levels of students’ understanding.</p> <p>5.7. assists students’ learning throughout the conversation by questioning, restating, praising, encouraging, etc.</p> <p>5.8. guides the students to prepare a product that indicates the Instructional Conversation’s goal was achieved.</p>
<b>Te Kotahitagan</b>
3. <i>Nga whakapiringatanga</i> : Teachers are able to create a secure, well-managed learning environment.
5. <i>Ako</i> : Teachers can use strategies that promote effective teaching interactions and relationships with their learners
4. <i>Wananga</i> : Teacher are able to engage in effective teaching interactions with Maori students as Maori
<b>TLRP</b>
6. Effective pedagogy promotes the active engagement of the learner: This involves acquiring a repertoire of learning strategies and practices, developing positive learning dispositions, and having the will and confidence to become independent agents in their own learning. The promotion of learner independence and autonomy, in this sense, is not just about the effectiveness of learning. It also concerns the realisation of rights, formation as a person, manifestation of citizenship and contribution of individuals to history.
7. Effective pedagogy fosters both individual and social processes and outcomes: Learners should be encouraged and helped to build relationships and communication with others for learning purposes, in order to assist the mutual construction of knowledge and enhance the achievements of individuals and groups.



*Facet Four*

<b>Facet IV. Taking an inquiry stance through using evidence to scaffold learning and improve teaching</b>
<b>MET FFT</b>
3d. Using assessment in instruction as the assessment for learning and monitoring students' understanding. This can be demonstrated through setting clear assessment criteria, monitoring students learning carefully in a way that is integrated well into lesson, providing timely/constructive/substantive feedback to guide/advance students' learning, creating clear criteria that promote students' self-assessment/monitoring of their own progress. Indicators include that the teacher paying close attention to evidence of student understanding; the teacher posing specifically created questions to elicit evidence of student understanding; the teacher circulating to monitor student learning and to offer feedback; students assessing their own work against established criteria.
3e. Demonstrating flexibility and responsiveness refer to a teacher's skill in making adjustments in a lesson to respond to changing conditions. This subdomain include elements like making minor/(at times) major adjustment to a lesson when needed, responding to students and seizing the "teachable moment" during the class, being persistent in seeking alternate approaches to help students learn. Indicators include incorporation of students' interests and daily events into a lesson; the teacher adjusting instruction in response to evidence of student understanding (or lack of it); the teacher seizing on a teachable moment.
<b>MET CLASS</b>
Domain 3. Instructional Support: dimension 2. Quality of feedback
<b>MET TRIPOD</b>
Clarify: Cultivate understanding and overcome confusion
<b>BES – Teaching Diverse Students</b>
1. Quality teaching is focused on raising student achievement (including social outcomes) and facilitate high standards of student outcomes for diverse learners. Quality teaching should facilitate the learning of diverse students and raises achievement for all learners. Teachers need to hold high expectations to all learners and recognize that the idea of learning styles approach has led to inappropriate stereotyping of minority students by teachers (e.g. minority students as kinaesthetic learners). Research suggests that the narrower, procedural hands-on approach was less effective in supporting student learning than a conceptual and child-centered approach that highlights children's thinking through interaction with equipment and emphasises the interrelationships between mathematical ideas. That is, inappropriate assumptions about ethnically-based learner needs can maintain the least effective learning opportunities for these students. To achieve the goal of raising achievement for all students, teachers establish and follow through on appropriate expectations for learning outcomes (e.g., through properly designed diagnostic assessment to facilitate learning) and the pace at which learning should proceed. Evidence suggests that while teacher expectations for high standards are necessary, expectations alone are insufficient to facilitate achievement when not supported by quality teaching. High expectations need to be integrated into quality teaching practices and demonstrates clear links between

pedagogical practices and achievement outcomes. Also, attention should be paid to the development and precision in the language and practice of diagnostic assessment. Diagnostic or formative assessment can play an integral role in assisting the teacher to raise student achievement as long as the assessment practices are integrally embedded within, and facilitative of quality teaching, rather than an alternative focus.

4. Quality Teaching is Responsive to Student Learning Processes for all students (particularly students with special needs) and in all curriculum areas: Quality teaching is optimized when teachers have a good understanding of, and are responsive to, the student learning processes involved. Teachers have knowledge of the nature of student learning processes in the curriculum area, can interpret student behavior in the light of this knowledge and are responsive, creative and effective in facilitating learning processes. The interactive teaching approach can involve these stages in the teaching process: 1) Preparation (teachers give an overview and background of the topic); 2) Exploration (clarify the topic and carry out exploration activity); 3) Children's question (assist children to generate questions on their own); 4) Specific investigation (planning and conducting the investigation); 5) Reflection (report the finding and evaluate the investigation process). The interactive teaching approach explicitly attended to both the social and cognitive aspects of the teaching-learning process and needs of diverse learners. In the case of teaching literacy - teachers pay attention to the developmental progressions of emergent literacy, the kinds of literacy practices that children engage at home and how their home literacy practice interact with the classroom literacy. Teachers must ensure sufficiently frequent and paced opportunity to interact with new content to allow learning to occur. For students with special needs, diagnosis is the key part of interactive teaching and a critical part of the success of such adaptive education approaches is the learner's access to well-structured peer and social interaction opportunities.

8. Pedagogy scaffolds and provides appropriate feedback on students' task engagement: Tasks and classroom interactions provide scaffolds to facilitate student learning (the teacher provides whatever assistance diverse students need to enable them to engage in learning activities productively, for example, teacher use of prompts, questions, and appropriate resources including social resources); Teaching develops all students' information skills and ensures students' ready access to resources when needed to assist the learning process; Students receive effective, specific, appropriately frequent, positive and responsive feedback. Feedback must be neither too infrequent so that a student does not receive appropriate feedback nor too frequent so that the learning process is subverted.

9. Pedagogy promotes learning orientations, student self-regulation, metacognitive strategies and thoughtful student discourse: Quality teaching promotes learning orientations, student self-regulation, and critical thinking; Teaching scaffolds reciprocal or alternating tuakana teina<sup>2</sup> roles in student group, or interactive work; Teaching promotes sustained thoughtfulness (e.g. through questioning approaches, wait-time, and the provision of opportunities for application and invention); Teaching makes transparent to students the links between strategic effort and accomplishment.

10. Teachers and students engage constructively in goal-oriented assessment: Teachers ensure that their assessment practices impact positively on students' motivation and improve their learning; Teachers and students have clear information about learning outcomes; Students have a strong sense of involvement in the process of setting specific learning goals; Pedagogy scaffolds and provides appropriate feedback on students' task engagement; Teachers manage the evaluative climate, particularly in context of public discussion, so that student covert or overt participation is supported, scaffolded and challenged without students being humiliated (academic norms are not undermined but supported by social norms); Teachers adjust their teaching to take account of the results of assessment.

**BES – Social Studies**

2.2 Aligning experiences through purposefully aligning activities and resources (i.e., spoken, written, visual, and experiential) to intended outcomes and the alignment must be continuously checked against the intended outcomes: Practices that purposefully align activities to desired outcome include: 1) expected understanding is explicitly included in the learning experiences; 2) expected understanding is implicitly included or partially embedded in the learning experiences; 3) the learning experiences incorporates additional information, explanation, and examples; 4) preparatory activity or discussion, and contextual information describing the focus or purpose of the activity are included. According to studies cited in the BES, these strategies seem to be consistently effective across different learning outcomes - use of strategies, development of conceptual knowledge, attitude and behavior, and social skills. Research findings also suggest that it may be necessary to align not only the instruction and the intended outcome but also the instruction and learner prior knowledge. Resources (e.g., spoken, written, visual, and experiential) should also be aligned to the purpose of a task that supports students in achieving outcomes related to that task. Teachers can also model the intended outcomes which makes alignment transparent to students, make the purpose of tasks clear to students (explain how the task will be assessed and provide resources such as templates, planning/writing guides, and structured overviews that scaffold them through the task) reducing ambiguity and helps them to focus on important learning. Teacher should also align assessment with teaching which helps communicate what is important.

**CREDE**

**4. Challenging Activities: Teaching Complex Thinking — Challenge students toward cognitive complexity:**

- 4.1. assures that students — for each instructional topic — see the whole picture as a basis for understanding the parts.
- 4.2. presents challenging standards for student performance.
- 4.3. designs instructional tasks that advance student understanding to more complex levels.
- 4.4. assists students to accomplish more complex understanding by building from their previous success.
- 4.5. gives clear, direct feedback about how student performance compares with the challenging standards.

**Te Kotahitagan**

6. *Kotahitanga*: Teachers promote, monitor and reflect on outcomes that in turn lead to improvements in educational achievement for Maori students

**TLRP**

4. Effective pedagogy requires learning to be scaffolded: Teachers, trainers and all those, including peers, who support the learning of others, should provide activities, cultures and structures of intellectual, social and emotional support to help learners to move forward in their learning.

5. Effective pedagogy needs assessment to be congruent with learning: Assessment should be designed and implemented with the goal of achieving maximum validity both in terms of learning outcomes and learning processes.

<p><b>Facet V. Taking an inquiry stance for further professional engagement and learning</b></p>
<p><b>BES Teaching Social Studies</b></p>
<p>5. Effective pedagogy requires teachers to <b>inquire into their priorities for learning, into interventions that might enhance student outcomes, and into the impact of their actions on their students.</b> Such inquiry increases the likelihood of student success and of teachers making a bigger difference. Teaching strategies may not work in all contexts: a strategy may work in one context but not another, for one student / group of students but not another, or in relation to one outcome but not another. Effective pedagogy requires teachers to inquire into: the outcomes that should be prioritized for their students (the focusing inquiry), teaching actions that are most likely to enhance outcomes for their students (the teaching inquiry), and into the actual impact of their actions on their students (the learning inquiry). <u>The focusing inquiry:</u> The focusing inquiry helps determine direction. Given that time is limited and that students need multiple opportunities to engage with the content of new learning, it is important to establish priorities; this is the purpose of this phase of the cycle. <u>The focusing inquiry:</u> The focusing inquiry helps determine direction. Given that time is limited and that students need multiple opportunities to engage with the content of new learning, it is important to establish priorities; this is the purpose of this phase of the cycle. <b>The focusing inquiry is termed an inquiry because the process of setting priorities draws from a variety of sources: curriculum requirements, community expectations, teacher interests and, most importantly, the learning needs, strengths and experiences of the learners.</b> This is not to say that student and parent expectations about priorities determine what is taught, but that those expectations inform teachers’ considerations about priorities alongside curriculum statements. <u>The teaching inquiry:</u> The teaching inquiry involves identifying strategies that are most likely to help the students achieve the prioritized outcomes. To do this, teachers must not <b>only locate evidence of effective strategies, but also evaluate its quality. Information is available from informal sources such as the teacher’s own experience as teacher and learner and the experiences of colleagues, prescriptive sources such as curriculum documents and textbooks, and systematic sources such as professional development and research.</b> <u>The learning inquiry:</u> The focus of the learning inquiry is the impact of the teacher’s actions on student learning and the implications for future teaching. This inquiry approach <b>requires that teachers be responsive to their students and their learning, to their own learning communities, to research and, in particular, to outcomes-linked evidence.</b></p>
<p><b>TLRP</b></p>
<p><b>OVERALL DESCRIPTION:</b> A distinctive characteristic of TLRP schools projects was their aim to generate new knowledge about effective teaching and learning in authentic settings, i.e. in classrooms led by teachers. In almost all cases this encouraged them to work directly with teachers, or other education professionals in classrooms, to develop innovations. This contrasts with much existing research on ‘what works’, especially from the United States, which tends to rely on university-</p>

based researchers to develop and test interventions in quasi-experimental settings. Under this system, those programmes, projects and products that achieve respectable effect sizes are disseminated through, for example, the What Works Clearinghouse. However, the transformation of evidence-based knowledge into sustainable and effective practice cannot be taken for granted. Promising innovations often fail simply because they are not implemented; and implementation depends on those who work on a daily basis with pupils taking ownership of new ideas and practices. This requires **teacher learning**. The general conclusion, to be drawn from the diverse studies in TLRP, is that changes in behavior and beliefs are both necessary and should be developed together, progressively. **Furthermore, effective pedagogy depends not only on behavioral change and the acquisition of new knowledge but on the development of values and dispositions, and reappraisal of roles and relationships in and beyond the classroom. Such learning by teachers takes place in the workplace, through participation in collaborative activities with other 'insiders', although the involvement of outsiders, such as researchers, and the provision of well-researched materials can be highly valued.**

9. Effective pedagogy depends on the learning of all those who support the learning of others: The need for lecturers, teachers, trainers and co-workers to learn continuously in order to develop their knowledge and skill, and adapt and develop their roles, especially **through practice-based inquiry**, should be recognised and supported. 1) **Learning is both individual and collective** and involves both the acquisition of knowledge and skills and participation in social processes. Thus the development of supportive professional cultures within which teachers can learn is vitally important. **Dynamic and expansive learning environments need to provide opportunities for boundary crossings, which encourage teachers to learn from others in different networks or communities of practice.** 2) Teachers are most ready to accept ideas for change if they resonate with their existing or previous beliefs and experience. However, **this does not make them right or appropriate. Teachers need to develop the knowledge and skills to evaluate evidence and the confidence to challenge taken-for-granted assumptions, including their own.** 3) **"...in order for teachers to change their practice a number of conditions are necessary: teachers need to believe that pupil perspectives are important; teachers need to help pupils to learn to take on the new roles demanded by consultation;** teachers need to be confident that they can combine responding to the needs of government and to the views of pupils; teachers need support from school managers to develop regular ways of consulting pupils. 4) **Collaborative learning: Teachers work in groups to formulate hypotheses about adjustments to lessons to improve learning.** These are tested in Research Lessons that colleagues observe and discuss subsequently. New hypotheses and adjustments are tested in further iterations until the teachers feel ready to perform a public research lesson. 5) **The purpose of practitioner research is to improve the ethical quality of teacher's interactions with students in the teaching-learning process rather than to drive up the standards in the classrooms.** The transformation of the culture of teaching and learning that prevailed in the field of humanities education, and which it is believed to be the primary source of students' disaffection, depends upon the capacity of teachers to adopt a research stance towards their practice. **This capacity**

is not considered to be in purely individualistic terms. Cultural transformation depends on teachers collaborating together across classrooms and schools to identify and diagnose common problems they experience in attempting to realize the standards implied by the pedagogical aim of developing understanding – given that their practice tends to be shaped by shared norms – and to devise experimental strategies for resolving them. **6) Teachers' resilience, commitment, and sense of positive professional identity:** Teachers demonstrated a capacity for strategic and reflective thinking and took responsibility for what happened in their classrooms. They were not inclined to blame external circumstances or pupil characteristics but concentrated on the ways in which they could improve the learning experiences for pupils. Teachers should accept their responsibility, ability, and power to make a crucial difference. The learning of teachers shares much with the learning of their students. All require a sense of purpose, a developed capacity for reflection and strategic thinking informed by evidence, motivation and a sense of their own agency to bring about improvements in outcomes.

10. Effective pedagogy demands consistent policy frameworks with support for learning as their primary focus: Organizational and system level policies need to recognise the fundamental importance of continual learning - for individual, team, organizational and system success - and be designed to create effective learning environments for all learners. 1) Support and commitment from administrator/leader/school senior management; 2) **advocacy: What is important is that all those with an interest in effective pedagogy – pupils, parents, teachers, researchers, policy makers and the public at large –strive together to find and establish socially just policy frameworks that truly support learning for diverse learners.**

<b>Facet VI. Challenging and Recognizing inequities</b>
<b>Te Kotahitagan</b>
One of assumptions underneath Te Kotahitanga is that culturally appropriate and responsive teachers should positively <b>reject deficit theorizing as a means of explaining Māori students' educational achievement.</b>
The purpose of TK is to improve the educational achievement of Maori students through <b>operationalizing Maori people's cultural aspirations for self-determination within non-dominating relations of interdependence</b> through developing classroom relations and interactions and in school institutions for this purpose. (Bishop et al., 2009, pp. 735)
The attempt to reduce <b>disparities does not just focus on bringing low achieving students up to the current levels of their peers</b> by traditional means; rather all students' achievement level needs to raise in order that educators can create learning contexts that will provide students with those tools that are vital for the future, the tools of creative, critically reflective thinking citizens. In order to do so, we need to <b>immerse students in power-sharing relationships with their peers and their teachers from an early age. In short, the principle of self-determination within nondominating relations of interdependence should be relevant to all involved in classroom interactions</b> (including teachers of course), and should raise educational achievement of all involved, whilst reducing disparities. (TK Professional Development)
<b>Maori's cultural ways of knowing, generating knowledge, and sense-making should be the centre of the "solution"</b> to Maori students' educational achievement and disparities rather than "a framework of neo/colonialism" which continues to serve the interests of a mono-cultural elite. (Bishop et al., 2009, pp. 735)
Operationalizing a Culturally Responsive Pedagogy of Relations involves implementing the Effective Teaching Profile (ETP). Such a profile creates a <b>learning context that is responsive to the culture of the child and means that learners can bring who they are to the classroom in complete safety and where their knowledge is acceptable and legitimate.</b> CRPP is an approach that rests in the first instance upon a commitment by teachers to build caring and learning relationships and interactions with Māori students; in the second, for teachers to strongly believe Māori students can improve their achievement; and thirdly, their students are able to take responsibility for their learning and performance.
Such a context for learning stands in contrast to the traditional classroom where the culture of the teacher is given central focus and has the power to define what constitutes appropriate and acceptable knowledge, approaches to learning, understandings and sense-making processes. This model suggests that <b>when the learner's own culture is central to their learning activities, they are able to make meaning of new information and ideas by building on their own prior cultural experiences and understandings.</b> The visible culture of the child need not necessarily be present but may well become



<p>present as a result of their <b>co-constructing learning experience with their teachers</b>. Such collaborative efforts address the potential imposition of the teacher displaying cultural iconography of their own choice. This shift from traditional classrooms is important because traditional classroom interaction patterns do not allow teachers to create learning contexts where the culture of the child can be present but rather assumes cultural homogeneity (Villegas &amp; Lucas, 2002), which in reality is cultural hegemony (Gay, 2000). (pp. 741)</p>
<p>This new pedagogy recognises that all people who are involved in the learning and teaching process are participants who have meaningful experiences, valid concerns, and legitimate questions. The teacher interacts with students in such a way (storying and re-storying) that <b>new knowledge is co-created</b>. Such a classroom will generate totally different interaction patterns and educational outcomes from a classroom where knowledge is seen as something that the teacher makes sense of and then passes onto students and will be conducted within and through a pedagogy of relations, wherein self-determining individuals interact with one another within non-dominating relations of interdependence.</p>
<p>Creating culturally responsive learning context where <b>the power of decision making and constructing knowledge in the classroom is shared</b>: teachers are able to create culturally appropriate and responsive contexts for learning in their classrooms in a manner where Maori students are able to interact with teachers and others in ways that <b>legitimizes who they are and how they make sense of the world</b>. (Bishop et al., 2009, pp. 736)</p>
<p>... we suggested that this (improving achievement for Maori students and all students) will be accomplished when educators <b>create learning contexts within their classroom; where power is shared between self-determining individuals within nondominating relations of interdependence</b>; where <b>culture counts</b>; where <b>learning is interactive, dialogic and spirals</b>; where participants are connected to one another through the establishment of a common vision for what constitutes excellence in educational outcomes. (The development of TK document; Bishop et al., 2009, pp. 736)</p>
<p>A change from the traditional transmission type of classrooms with monitoring and behavioral feedback to one with a wide range of interactions between the teacher and students. <b>These interactions include some instruction (a mixture of process and transmission), the monitoring of processes and uptake, the recognition of appropriate student behavior and in addition, the teachers increasingly acknowledge students' prior learning and respond to student-initiated interactions by giving academic feedback and feed-forward. They also co-construct the content and process of learning with students as co-learners.</b> Further, as teachers move towards a more discursive classroom, <b>they spend less time interacting with the whole class and more time with individuals and/or groups</b>. (Bishop et al., 2009, pp. 740)</p>
<p>Another instructional change is from the transmission of product/content focus to process-oriented instruction that promotes cooperative learning. Monitoring also changes from testing for compliance, content reception or understanding of</p>

instructions to monitoring of learning processes, which again can be subdivided into **monitoring of facilitated learning experiences** or the **monitoring that occurs during co-construction sessions**. In addition in the traditional classroom, feedback is provided on behavior as much as it is provided on academic initiatives, and both forms of feedback are limited. In addition, when the classes are at their most traditional, teachers will often provide behavioral feedback, “good boy, good girl” to an inquiry or an answer from a student that should receive an academic response. As the classes become more discursive, academic feedback increases markedly and behavioral feedback diminishes. (Bishop et al., 2009, pp. 740)

**Anti-deficit thinking and agentic positioning - Reflect, recognize, and challenge deficit thinking:** The teacher **recognizes and challenges their own deficit thinking and theorizing Maori students, and engages in critical reflection on their own discursive positioning and the impact to students**. When teachers are able to engage in critical reflection about the images they have of marginalized students and the resultant relationships they have with these students, **they are more likely to be able to engage in power-sharing practices** (i.e., seeing students as knowledgeable participants). The teacher explicitly rejects deficit theorizing as a means of explaining Maori students’ educational achievement levels, and **their taking an agentic position in their theorizing about their practice; that is, practitioners expressing their professional commitment and responsibility to bringing about change in Maori students’ educational achievement by accepting professional responsibility for the learning of their students**. It is an approach that rests in the first instance upon a commitment by teachers to build **caring and learning relationships and interactions with Maori students**; in the second, for teachers **to strongly believe Maori students can improve their achievement; and thirdly, their students are able to take responsibility for their learning and performance**. (Bishop et al., 2009, pp. 735-736; TK Student Interview)

The teacher believed they have a high level of understanding of the importance of relating to Maori students from an agentic position and in ensuring that their **teaching practices reflect an agentic attitude towards these target students**. (Bishop et al., 2009, pp. 739)

Fundamental to the ETP is teachers’ understanding the need to **explicitly reject deficit theorizing as a means of explaining Māori students’ educational achievement levels**, and their **taking an agentic position in their theorizing about their practice**. That is, practitioners expressing their professional commitment and responsibility to bringing about change in Māori students’ educational achievement by **accepting professional responsibility for the learning of their students**. (The development of TK document)

**Manifestation of ETP in classrooms:** the teachers demonstrate on a daily basis that they care for the students as Maori; they have high expectations of the learning for students; they are able to manage their classrooms so as to promote learning; they are able to engage in a range of discursive learning interactions with students or facilitate students to engage with others in these ways; they know a range of strategies that can facilitate learning interactions; they promote, monitor and reflect upon student learning outcomes that in turn lead to changes in teachers’ practice

that will bring about improvements in Maori student achievement, and teachers share this knowledge with their students. (Bishop et al., 2009; The development of TK)

Good relationships are based on teachers embracing all aspects of the ETP, including caring for students as culturally-located individuals as Maori, caring for their performance, and using a wide range of classroom interactions, strategies, and outcome indicators to inform their practice. (TK Student interview)

## Appendix C: General Descriptions for Six Facets (Version One)

### Facet One

#### **Facet One: Selecting worthwhile content and designing and implementing learning opportunities aligned to valued outcomes**

**Characteristic 1: Set clear and high learning goals and outcomes:** Teachers set clear and high standards of learning goals/outcomes, make the learning goals transparent to students, and constantly hold students toward the high expectations (*Standards Drawn: MET FFT 3a; MET TRIPOD Challenge; BES Teaching Diverse Students 1; CREDE 4.2; TK 2; TLRP 1*)

- ✓ Set high standards of learning goals and outcomes encompassing aspects of intellectual, socio-emotional, critical and civic learning and development. (*MET TRIPOD Challenge;*
- ✓ Communicate clearly these high standards and important educational values of learning outcomes with students.
- ✓ Follow through high expectation towards all students and work to raise achievement for all.

**Characteristic 2: Design and select challenging learning opportunities:** Teachers select and design learning experiences that support students' understanding of content and conceptual development, advance their complex/high-cognitive learning, and facilitate students' active participation in the learning process (*Standards Drawn: MET FFT 3a, 3b, 3c; MET CLASS Domain 3; MET TRIPOD Challenge, Consolidate; BES Teaching Diverse Students 5; BES Social Studies, 2.2, 2.3; CREDE 4; TLRP 2*)

- ✓ Teachers recognize and demonstrate their subject content and pedagogical knowledge, and reflect on their own biases in designing and selecting learning opportunities (*MET FFT 3b; BES Social Science 2.1*)
- ✓ Explain instructional concepts and directions to students by using clear, vivid, precise, and accurate language as well as using a variety of communication strategies such as metaphor and analogies. (*MET FFT 3a; MET CLASS Domain 3*)
- ✓ Select and design a wide range of learning experiences (i.e., questions, discussion technique, tasks/assignments) to support students' conceptual and content-based learning, engage students in developing deep understanding, critical thinking, reflection and problem solving, as well as facilitate/empower students' participation and expression of their own learning (*MET FFT 3b, 3c; MET CLASS Domain 3; MET TRIPOD Consolidate; BES Teaching Diverse Students 5; CREDE 4.1, 4.3, 4.4*)

**Characteristic 3: Align learning experiences with prior knowledge and desired outcomes:** Teachers make the connection between learning experiences and desired outcome transparent to students by providing sufficient and effective learning opportunities. (*Standards Drawn: BES Teaching Diverse Students 5, 6, 7; BES Teaching Social Studies 2.1~4*)

- ✓ Align learning experiences with desired outcomes by employing a wide range of techniques to understand the learning needs and prior knowledge of diverse learners and address diversity (*BES Teaching Diverse Students 5; BES Social Studies 2.1, 2.4*)
- ✓ Purposefully align activities and resources to intended outcomes and make the alignment between learning experiences and the learning outcomes/purposes clear to students. (*BES Teaching Social Studies 2.2; BES Teaching Diverse Students 7*)
- ✓ Provide sufficient (i.e., multiple and frequent) and effective (i.e., carefully, structured, intentional repetition of the same information, questions to facilitate personalization of knowledge) learning opportunities to support the learning and development needs of students. (*BES Teaching Diverse Students 6; BES Teaching Social Studies 2.3*)

*Facet Two*

**Facet Two: Connecting to students as learners and their lives and experiences**

**Characteristic 1. Identify and recognize students' home and community culture**

- ✓ Identify and recognize students' prior knowledge, experiences, home and community culture by using a variety of approaches (i.e., listening, talking, writing activities, etc.). (*MET CLASS Domain1; MET TRIPOD Care; BES Teaching Diverse Students 3; CREDE 2, 3; TK 1; TLRP 3*)
- ✓ Recognize the strengths within the families/homes of diverse students and proactively cultivate the partnerships with parents/caregivers. (*BES Diverse Students 3; CREDE 3*)

**Characteristic 2. Make connection between school culture context and students' home/community cultural contexts**

- ✓ Use aboriginal culture philosophy as the underpinning of curriculum design rather than “add-on.” (*BES Social Studies 1.1; CREDE 2, 3; TK 5; TLRP 3*)
- ✓ Respond to students' prior knowledge, skills, diverse interests, motivations and level of engagement by making curriculum content relevant and inclusive to students' lives. Specifically, teachers need to ensure that the students' experience of instruction is relevant and meaningful to their real life and this connection must be made explicit and transparent to students. Teachers also need to design a variety of instructional activities (e.g., project based learning initiated by students, real-world situations and experiences, informal learning) to meet students' diverse motivation, interests. (*MET TRIPOD Captive; BES Teaching Diverse Students 3; BES Social Studies 1.1, 1.2, 4.1, 4.2; CREDE 3; TLRP 8*)
- ✓ Utilize students' culture as pedagogical resources, embed students' cultural knowledge and experiences in the content, and use cultural resources as counter stories to encourage students to challenge and criticize normalized and privileged discourse in the mainstream textbook. (*BES Teaching Diverse Students 3; BES Social Studies 1.1*)
- ✓ Teachers' use of language and communication approaches puts diversity front and center. Teachers also encourage students to use content vocabulary to express their understanding by using a variety of modes of expression (e.g., using both first and second languages, writing, activities). (*BES Social Studies, 1.2; CREDE 2, 3*)

*Facet Three*

**Facet Three: Creating learning-focused, respectful and supportive learning environments**

**Characteristic 1: Foster a caring, respectful, and inclusive learning environment**

*(MET FFT 2a, 2d; MET CLASS Domain 1, 2; MET TRIPOD Control; BES Diverse students 2.1, 2.2, 2.3, 2.4, 2.6, 2.7, 2.9; BES Social Studies 3.1, 3.2; CREDE 5; TK 3, 4, 5; TLRP 7)*

- ✓ Teachers establish cognitive, social, emotional, and spatial connections with students. Teachers show warmth, caring, encouragement, and respect in interacting with students (e.g., active listening, giving feedback) and value/recognize the diversity of students, and support students to express their views. The reciprocal relationships between teachers and students create a sense of belongingness to the school and classroom community and increase students' motivation to learn. *(MET FFT 2a; MET CLASS Domain 1; BES Diverse students 2.1, 2.2, 2.4, 2.6; BES Social studies 3.1; CREDE 5; TK 3, 4)*
- ✓ Teachers set a clear expectation of students' classroom conduct and group working norm, and respond to students' behaviors in a way that respect students' dignity and reinforce positive/respectful behaviors. For example, teachers can provide explicit instructions to foster students' skills on listening, questioning, speaking, seeking help, acknowledging others, modeling cooperative learning, and acknowledging multiple abilities and voices. Interactions between teachers and students and among students show mutual respect, trust, and effective collaboration *(MET FFT 2d; MET CLASS Domain 1, 2; MET TRIPOD Control; BES Diverse Students 2.1, BES Social Studies 3.2)*

**Characteristic 2: Use instructional practice that establishes a learning-focused and collaborative learning community**

*(MET FFT 2a, 2b; MET TRIPOD Control, Confer; BES Diverse students 2.1, 2.3, 2.4, 2.5, 2.8, 2.9, 2.10, 2.11; BES Social Studies 3.2, 3.3; CREDE 1, 5; TLRP 6, 7)*

- ✓ Teachers set high expectation of learning for all students, convey the educational value of content of learning and the academic norm of classroom, and acknowledge the high-quality work and persistence on the part of students. *(MET FFT 2b; BES Diverse students 2.4, 2.11; CREDE 5)*
- ✓ Teachers use effective pedagogical practices to facilitate and support students' collaborative learning culture and independent/persistent learning dispositions by sharing power and responsibility for decision-making with students. For example, teachers design complex tasks or multi-tasks environment that require students' cooperative teamwork and learning and discussion with each other. The designed instructional tasks/format also value and require the diverse experiences of students as a resource and strength for their learning. *(MET FFT 2a, 2b; MET TRIPOD Confer; BES Diverse students 2.1, 2.3, 2.5, 2.8, 2.9, 2.10, 2.11; BES Social Studies 3.2, 3.3; CREDE 1, 5; TLRP 6, 7)*

**Characteristic 3: Manage classroom procedures and physical space to facilitate collaborative learning** (*MET FFT 2c, 2e; MET CLASS Domain 2; BES Diverse Students 2.1; CREDE 1, 5*)

- ✓ Teachers manage classroom procedures efficiently, effectively, and purposefully so that learning time is maximized. This includes preparation and management of instructional materials, clear instructions provided to students so that they know where to move and what to do (*MET EFF 2c; MET CLASS Domain 2; BES Diverse Students 2.1; CREDE 1*)
- ✓ Teachers organize the classroom physical space to be safe and accessible to all students and arrange the furniture/facilities (e.g., technology) in a way that maximizes students' learning. For example, teachers use certain group seating plan to accommodate the need for joint group communication or work. (*MET FFT 2e, CREDE 1, 5*)



*Facet Four*

**Facet Four: Taking an inquiry stance through using evidence to scaffold learning and improve teaching**

**Characteristic 1: Design and integrate assessment into curriculum and instruction to scaffold students' learning:** Teachers must pay close attention to the evidence of students' learning, use and integrate assessment into classroom instructions to monitor students' progress, respond to their needs, and improve instructions. (*MET FFT 3d, 3e; MET TRIPOD Clarify; MET CLASS Domain 3; BES Teaching Diverse Students 1, 4, 8, 9, 10; BES Social Studies 2.2; CREDE 4; TK 6, TLRP 4, 5*)

- ✓ The teacher designs the classroom assessment that has clear criteria and is integrally embedded within instructional activities (*MET FFT 3d; BES Teaching Diverse Students 1, 10; BES Social Studies 2.2; TLRP 5*)
- ✓ To monitor students' learning and provide timely, substantive, and constructive feedback to guide and advance students' learning, the teacher circulates and interacts with students. The teacher also uses a variety of approaches (e.g., portfolio, artifacts, as well as traditional tests) to assess students' learning, including posing specifically created questions or using prompts to elicit evidence of diverse students' learning. (*MET FFT 3d; MET CLASS Domain 3; BES Teaching Diverse Students 8; CREDE 4; TLRP 4*)
- ✓ The teacher involves students in the process of setting specific learning goals, and engages constructively in goal-oriented assessment. The teacher carefully manages the evaluative climate and ensures the positive impact of assessment practice on students' motivation and learning. The teacher's use of assessment practice also nurtures students' learning orientation, self-regulation, critical thinking, and metacognitive strategies. (*BES Teaching Diverse Students 9, 10*)
- ✓ The teacher uses results of diagnostic assessment to adjust teaching practice. The teacher demonstrates his/her flexibility and responsiveness to adjust instruction (e.g., seize the "teachable moment") and seeks alternate approaches to help students to learn. (*MET FFT 3e; MET TRIPOD Clarify; BES Teaching Diverse Students 4, 10; TK 6*)

**Facet Five: Taking an inquiry stance for further professional engagement and learning**

**Characteristic 1: Takes an inquiry stance for individual and collective learning**

Teachers are reflective of their practices and their own biases/assumptions and are receptive of changes that do not resonate with their existing values, beliefs, and experiences. Specifically, teachers continuously learn to develop their knowledge and skills and adapt their roles through practice-based inquiry into the priorities for student learning, their instructional interventions, and the impact of their actions on students' learning outcomes (*BES Social Studies, TLRP 9*).

- ✓ Teachers as practitioner researchers inquire into their priorities for student learning by drawing a variety of sources and engaging with various stakeholders, including curriculum requirements, community expectations, teacher interests and, most importantly, the learning needs, strengths and experiences of the learners. Teachers believe and trust the importance perspectives of learners and facilitate learners to take on the active role in their own learning.
- ✓ Teachers as practitioner researchers inquire into their instructional interventions that might enhance student learning. Teachers work in groups to identify the evidence of effective strategies from a variety of sources (including teaching experiences, curriculum documents and textbooks, professional development, and research) and post hypotheses on their actions and adjustments made.
- ✓ Teachers monitor and evaluate the impacts of their actions on their students' learning based on evidence. Teachers respond to their students and their learning, to their own learning communities, to research and, in particular, to outcomes-linked evidence.
- ✓ Teachers reflect on their own practice, subject content knowledge, biases, and assumptions. They are open for changes that might not reflect their experiences and assumptions, take responsibility to make a difference and to bring about improvement in student learning outcomes.

**Characteristic 2: Collaborates with involved stakeholders to advocate for supportive and sustainable learning environment and professional culture (*TLRP 9, 10*)**

- ✓ Teachers have strong commitment, resilience and sense of positive professional identity and take responsibility for what happened in their classrooms.
- ✓ Teachers collaborate with stakeholders - students, parents/caregivers, community members, colleagues, researchers, policy makers and the general public – to advocate for a sustainable and professional learning environment and policy frameworks that support the learning of diverse students.

### Facet Six: Recognizing and Challenging Inequities

#### **Characteristic 1: Recognize and challenge one's own deficit thinking and theorizing historically marginalized students** (*BES Teaching Diverse Students 3; BES Social Studies 1.1, TK*)

- ✓ The teacher rejects deficit theorizing as a means of explaining marginalized students' educational achievement and engages in critical reflection on their discursive positioning and the impact to students. Moreover, the teacher recognizes that the attempt to reduce education disparities cannot just focus on bringing low achievement students up to the current levels of their peers by traditional means; rather, all students' achievement must be raised so that they become creative and critical thinking citizens.
- ✓ The teacher regards the culture of marginalized students', their ways of knowing, generating knowledge, and sense-making as the source/center of his/her classroom practice and student learning experiences rather than assuming an elite mono-cultural framework. The teacher creates a learning context that is responsive and relevant to the culture of students, who they are and how they make sense of the world.
- ✓ The teacher utilizes students' culture as pedagogical resources, embed students' cultural knowledge and experiences in the content, and use cultural resources as counter stories to encourage students to challenge and criticize normalized and privileged discourse in the mainstream textbook.
- ✓ The teacher shares the power of decision-making and constructing knowledge with learners in the classroom. Rather than centering the culture of the teacher in the learning and teaching process, learners are able to make meaning of new information and ideas by building on their own cultural experiences and understandings. New knowledge is co-constructed through dialogic interactions among teacher and learners based on the principle of self-determination within the non-dominating relationship of interdependence.

#### **Characteristic 2: Take an agentic position in their practice and accept professional commitment and responsibility** (*TK*)

- ✓ The teacher expresses his/her professional commitment and responsibility to bringing about change in marginalized students' educational achievement by accepting professional responsibility for the learning of their students.
- ✓ The teacher believes that students (especially historically marginalized students) can improve their achievement and their students are able to take responsibility for their learning and performance.

## Appendix D: General Descriptions for Six Facets (Version Two)

### *Facet One*

#### **Facet One: Selecting worthwhile content and designing and implementing learning opportunities aligned to valued outcomes**

##### ***Characteristic 1: Select content and set learning goals and outcomes***

Teachers care and support for their students' performance. They set high standards of learning goals and outcomes, and make the learning goals and outcomes transparent, clear, and meaningful to students. Teachers follow through high expectation towards all students and focus on raising achievement for all.

##### ***Characteristic 2: Design and select learning opportunities***

Teachers demonstrate their subject content and pedagogical knowledge in instructional practices and reflect on their own biases. Teachers use a variety of communication approaches to reach out students and their use of language in instructional dialogue is clear, free of content errors, precise, and vivid. Teachers select and design a variety of instructional strategies and activities to support students' content-based and conceptual understanding. Teachers also select and design a variety of learning opportunities to motivate diverse learners and to maintain their intellectual engagement. Specifically, teachers use a wide range of activities, materials, and techniques (e.g., teacher-directed groups, cooperative work, individual work, structured peer interaction) challenge students to advance their thinking to the more complex and conceptual level rather than rote learning on procedural tasks, to apply big ideas for problem solving, and to critically reflect on one's own learning and thought process. Most importantly, teachers involve students in the process of actively examining, critiquing and constructing knowledge, and facilitate students' active participation in their own learning.

##### ***Characteristic 3: Implement planned learning experiences aligned with valued outcomes***

Teachers align learning experiences with desired outcomes by first identifying learning needs and prior knowledge of diverse students. By doing so, teachers are then able to align instructional activities and resources (e.g., ICT usage, task design, materials used) to intended outcomes, and make the alignment clear to students. Specifically, teachers explicitly or implicitly include the expected outcomes in the learning experiences, incorporate additional information, explanation, illustrations into the learning experiences, and provide the contextual information describing the purpose of learning activities. In addition, teachers provide sufficient (i.e., multiple and frequent) and effective (i.e., carefully, structured, intentional repetition of the same information, questions to facilitate personalization of knowledge) learning opportunities to support students' concept and skill development, so that task cycle matches students' developmental learning cycles.

*Facet Two*

**Facet Two: Connecting to students as learners and their lives and experiences**

***Characteristic 1. Identify and recognize students' home and community culture***

Teachers utilize a variety of strategies (e.g., listening/talking to students in oral or written forms) to identify students' prior knowledge, skills, experiences and home culture. Recognizing the strength within the homes/communities of diverse students, teachers initiate the effort to build relationships with parents/caregivers and to create opportunities for parents/caregivers to participate in classroom instructional activities.

***Characteristic 2. Make connection between school culture context and students' home/community cultural contexts***

Teachers use the philosophy of culture as the underpinning of the curriculum design rather than "add-on." To respond to students' prior knowledge, skills, interests and motivations, teachers use students' culture as pedagogical resources and embed students' cultural knowledge and experiences in the curriculum. To ensure the explicit, relevant, and meaningful connection between students' experience of instruction and their real life, teachers design a variety of instructional activities to respond to the diversity of students. Teachers use a variety of communication approaches to explain instructions and encourage students to express their content understanding by using a variety of modes of expression as well. Altogether, the premise, pedagogy and curriculum of teachers' instructional practices encourage students to challenge and criticize normalized and privileged discourse in the mainstream textbook.

### *Facet Three*

#### **Facet Three: Creating learning-focused, respectful and supportive learning environments**

##### ***Characteristic 1: Foster a caring, respectful, and inclusive learning environment***

The teacher shows warmth/physical proximity to their students, listens to diverse students' viewpoints attentively, values and recognizes the diversity of students, and provides feedback to support their students. The teacher has productive and reciprocal relationships with students, which make their students feel belonging to the community and motivated to learn. To create a positive classroom environment and cooperative peer culture, the teacher facilitates constructive behaviors among students by having a clear expectation toward classroom conduct and group working norm, providing explicit instructions to develop students skills on interacting/working with each other, and modeling the interactional skills.

##### ***Characteristic 2: Use instructional practice that establishes a learning-focused and collaborative learning community***

The teacher sets high expectation of learning for all students by conveying the educational values of content of learning and academic norm of classroom. The teacher designs complex tasks (open-ended, uncertain) that require collaborative knowledge construction, diverse experiences and perspectives, and productive dialogue among students. The teacher uses purposeful grouping of students (e.g., jigsaw learning groups – group members consist of students of different types of ability, interests, age, etc.) and selects task formats to foster the culture of peer learning and to fulfill instructional purposes. Most importantly, the teacher shares power and responsibility for decision-making with students in order to develop students' independent and persistent learning dispositions.

##### ***Characteristic 3: Manage classroom procedures and physical space to facilitate collaborative learning***

To accommodate instructional practices, the teacher prepares instructional materials/resources and manages the classroom procedures/logistics efficiently, effectively, and purposefully so that students' learning time is maximized. The teacher also arranges the classroom space, such as furniture and technology facilities, according to the need for group communication and joint work.

*Facet Four*

**Facet Four: Taking an inquiry stance through using evidence to scaffold learning and improve teaching**

***Characteristic 1: Design and integrate assessment into curriculum and instruction to scaffold students' learning***

The teacher designs the classroom assessment that is integrally embedded within instructional activities. To promote an evaluative climate that has positive impact on students' motivation and learning orientations, the teacher involves students in the process of setting specific learning goals and engages with students in constructive goal-oriented assessment. The teacher pays close attention to the evidence of students' learning through integrating a variety of diagnosis assessment approaches (e.g., portfolio, art facts, questions, prompts, as well as traditional tests) into instructions to monitor students' progress, adjust instructional design according to their needs, and scaffold students' learning to maximize their learning outcomes. Teachers set clear assessment criteria and communicate clearly to students without ambiguity. With the integrated assessment approaches, teachers provide timely, substantive, constructive, clear, and frequent-enough feedback to guide and advance students' learning. The teacher demonstrates his/her flexibility and responsiveness, such as adjusting their pedagogy according to the results of diagnostic assessment, seizing the "teachable moment," and seeking alternate approaches to help students to learn.

*Facet Five*

**Facet Five: Taking an inquiry stance for further professional engagement and learning**

***Characteristic 1: Takes an inquiry stance for individual and collective learning:***

Teachers are reflective of their practices, subject content knowledge, and their own biases and assumptions. They are willing to accept changes that do not resonate with their existing values, beliefs, and experiences by constantly involving in practice-based research with their colleagues. Specifically, teachers continuously learn to develop their knowledge and skills, and adapt their roles by inquiring into the priorities for student learning from a variety of sources including students' perspectives. Teachers also work together to identify the evidence of effective strategies based a variety of sources such as their experiences and research, post hypotheses on their instructional actions, monitor and evaluate the impacts of their actions on their students' learning, and most importantly respond to the findings.

***Characteristic 2: Collaborates with involved stakeholders to advocate for supportive and sustainable learning environment and professional culture***

Teachers have strong commitment, resilience and sense of positive professional identity and take responsibility for what happened in their classrooms. Teachers collaborate with stakeholders such as students, parents/caregivers, community members, colleagues, researchers, policy makers and the public at large to advocate for a sustainable and professional learning environment and policy frameworks that support the learning of diverse students.



**Facet Six: Recognizing and Challenging Inequities**

***Characteristic 1: Recognize and challenge one's own deficit thinking and theorizing historically marginalized students***

The teacher rejects deficit thinking and theorizing as a means of explaining marginalized students' educational achievement and engages in critical reflection on their discursive positioning and the impact to students. Rather than assuming an elite mono-cultural framework in the classroom, the teacher regards the culture of marginalized students, their ways of knowing, generating knowledge, and sense-making as the source/center of teaching practice and student learning experiences in the classroom. The teacher utilizes students' culture as pedagogical resources, embed students' cultural knowledge and experiences in the content, and use cultural resources as counter stories to encourage students to challenge and criticize normalized and privileged discourse in the mainstream textbook. The teacher shares the power of decision-making and constructing knowledge with learners in the classroom. New knowledge is co-constructed through dialogic interactions among teacher and learners based on the principle of self-determination within the non-dominating relationship of interdependence.

***Characteristic 2: Take an agentic position in their practice and accept professional commitment and responsibility***

The teacher accepts his/her professional commitment and responsibility to bring about change in marginalized students' educational achievement. He/she believes that students are able to take responsibility for their learning and can improve their achievement performance.

## Appendix E: Descriptions of Levels of Enactment for Six Facets

### Facet One

#### **Facet One: Selecting worthwhile content and designing and implementing learning opportunities aligned to valued outcomes**

<b>High Level of Enactment</b>	
<p>The teacher holds high expectations for all students by setting and communicating cognitively challenging and worthwhile learning goals. The teacher selects content that reflects valued forms of knowledge as expressed in the curriculum and by relevant communities. The teacher uses a range of instructional approaches and real world relevant resources to cultivate students' conceptual and critical learning and to empower students to take initiatives for their own learning. Instructional explanations are clear, relevant, error free, and meaningful to all students. The teacher skillfully and appropriately uses techniques, moves and strategies to provide rich and effective learning opportunities for all students.</p>	
<b>Critical Attributes</b>	<p><b>Characteristic 1: Selects content and sets learning goals and outcomes</b></p> <ul style="list-style-type: none"> <li>✓ Academic learning goals and outcomes are cognitively challenging and reflect valued knowledge in the curriculum and in communities.</li> <li>✓ Social, emotional, creative, and other learning goals are appropriately demanding and reflect valued knowledge in the curriculum and in communities.</li> <li>✓ The teacher clearly and consistently communicates high expectations, meaningful learning goals and relevant outcomes in order to empower students to take initiative for their own learning.</li> </ul> <p><b>Characteristic 2: Designs and selects learning opportunities</b></p> <ul style="list-style-type: none"> <li>✓ The teacher designs learning opportunities that are relevant to the curriculum and to the lives and experiences of students, as evidenced by students' desire, enthusiasm, and excitement to learn in classrooms.</li> <li>✓ Teacher purposefully selects a wide range of rich and relevant learning opportunities to cultivate conceptual and critical understanding as well as problem-solving ability of all students.</li> </ul> <p><b>Characteristic 3: Implements planned learning experiences aligned with valued outcomes</b></p> <ul style="list-style-type: none"> <li>✓ The teacher demonstrates strong content knowledge by providing clear, error-free and compelling explanations, demonstrations and exemplars of key concepts</li> <li>✓ The teacher appropriately uses a variety of communication approaches, including oral, written, and visual strategies, to capture students' interests build on their experiential, cultural and linguistic resources.</li> <li>✓ The teacher constantly monitors student learning and appropriately draws from a rich repertoire of strategies, moves and techniques to advance student performance aligned with high expectations.</li> </ul>
<b>Moderate Level of Enactment</b>	
<p>The teacher holds high expectations for some students by setting and communicating</p>	

<p>cognitively challenging and worthwhile learning goals. The teacher sometimes selects content that reflects valued forms of knowledge as expressed in the curriculum and by relevant communities. The teacher sometimes uses instructional approaches and real world relevant resources to cultivate students’ conceptual and critical learning and to empower students to take initiatives for their own learning. Instructional explanations are sometimes unclear, irrelevant, contain errors, and are not meaningful to all students. The teacher sometimes uses appropriate techniques, moves and strategies to provide effective learning opportunities for students.</p>	
<p><b>Critical Attributes</b></p>	<p><b>Characteristic 1: Selects content and sets learning goals and outcomes</b></p> <ul style="list-style-type: none"> <li>✓ Academic learning goals and outcomes are sometimes not cognitively challenging and do not always reflect valued knowledge in the curriculum and in communities.</li> <li>✓ Social, emotional, creative, and other learning goals are sometimes inappropriately demanding and do not always reflect valued knowledge in the curriculum and in communities.</li> <li>✓ The teacher is sometimes unclear and inconsistent in communicating high expectations, meaningful learning goals and relevant outcomes, which reduces opportunities for students to take initiative for their own learning.</li> </ul> <p><b>Characteristic 2: Designs and selects learning opportunities</b></p> <ul style="list-style-type: none"> <li>✓ Teacher sometimes designs learning opportunities that are not relevant to the curriculum and to the lives and experiences of students, as evidenced by students’ limited desire, enthusiasm, and excitement to learn in classrooms.</li> <li>✓ Teacher sometimes lacks purpose when selecting learning opportunities intended to cultivate conceptual and critical understanding as well as problem-solving ability of all students.</li> </ul> <p><b>Characteristic 3: Implements planned learning experiences aligned with valued outcomes</b></p> <ul style="list-style-type: none"> <li>✓ The teacher content knowledge is inconsistent. Sometimes explanations of key concepts are not clear or compelling, contain errors, and demonstrations and exemplars are not always relevant.</li> <li>✓ The teacher uses a limited range of communication approaches, including oral, written, and visual strategies, and does not consistently capture students’ interests by building on their experiential, cultural and linguistic resources.</li> <li>✓ The teacher inconstantly monitors student learning and draws from a limited repertoire of strategies, moves and techniques to advance student performance aligned with high expectations.</li> <li>✓</li> </ul>
<p><b>Low Level of Enactment</b></p>	
<p>The teacher holds high expectations for very few students and rarely sets and communicates cognitively challenging and worthwhile learning goals. The teacher seldom selects content that reflects valued forms of knowledge as expressed in the curriculum and by relevant communities. The teacher almost never uses instructional approaches and real world relevant resources to cultivate students’ conceptual and critical</p>	

<p>learning and does not empower students to take initiatives for their own learning. Instructional explanations are often unclear, irrelevant, contain errors, and are not meaningful to all students. The teacher rarely uses appropriate techniques, moves and strategies to provide effective learning opportunities for students.</p>	
<p><b>Critical Attributes</b></p>	<p><b>Characteristic 1: Selects content and sets learning goals and outcomes</b></p> <ul style="list-style-type: none"> <li>✓ Academic learning goals and outcomes are often not cognitively challenging and often do not reflect valued knowledge in the curriculum and in communities.</li> <li>✓ Social, emotional, creative, and other learning goals are often inappropriately demanding and often do not reflect valued knowledge in the curriculum and in communities.</li> <li>✓ The teacher is usually unclear and inconsistent in communicating high expectations, meaningful learning goals and relevant outcomes, which greatly reduces opportunities for students to take initiative for their own learning.</li> </ul> <p><b>Characteristic 2: Designs and selects learning opportunities</b></p> <ul style="list-style-type: none"> <li>✓ Teacher rarely designs learning opportunities that are relevant to the curriculum and to the lives and experiences of students, as evidenced by students' lack of desire, enthusiasm, and excitement to learn in classrooms.</li> <li>✓ Teacher frequently lacks purpose when selecting learning opportunities intended to cultivate conceptual and critical understanding as well as problem-solving ability of all students.</li> </ul> <p><b>Characteristic 3: Implements planned learning experiences aligned with valued outcomes</b></p> <ul style="list-style-type: none"> <li>✓ The teacher content knowledge is weak. Often explanations of key concepts are unclear, not compelling, contain errors, and demonstrations and exemplars are often not relevant.</li> <li>✓ The teacher uses a very limited range of communication approaches (including oral, written, and visual strategies) and does not capture students' interests by building on their experiential, cultural and linguistic resources.</li> <li>✓ The teacher rarely monitors student learning and draws from a very limited repertoire of strategies, moves and techniques so does not advance student performance aligned with high expectations.</li> </ul>

Facet Two

**Facet Two: Connecting to students as learners and their lives and experiences**

<b>High Level of Enactment</b>	
<p>The teacher identifies, recognizes and utilizes the students’ experiential, cultural and linguistic resources and sees all aspects of diversity as valued assets rather than as deficits. In designing activities and assessments the teacher draws on students’ prior knowledge, experiences, home and community cultures to facilitate learning that is highly relevant and meaningful to their lives, interests, and motivation. The teacher cultivates relationships with parents/caregivers and community members and includes them as partners in enhancing their children’s learning outcomes and opportunities.</p>	
<b>Critical Attributes</b>	<p><b>Characteristic 1. Identifies and recognizes students’ home and community culture</b></p> <ul style="list-style-type: none"> <li>✓ Almost all the time, the teacher recognizes the diverse experiential, cultural and linguistic resources students bring to school and uses a variety of approaches to incorporate these in the classroom and more broadly.</li> <li>✓ The teacher consistently recognizes the strengths of students’ home and community cultures, avoiding deficit and stereotypical thinking</li> <li>✓ The teacher builds positive relationships with parents/caregivers and includes them as partners in enhancing their children’s learning outcomes and opportunities.</li> </ul> <p><b>Characteristic 2. Makes connections between school and home cultures by planning and implementing linguistically and culturally sensitive curriculum, instruction and assessment</b></p> <ul style="list-style-type: none"> <li>✓ The teacher consistently embeds students’ cultural knowledge/experiences into learning experiences through deliberate selection of materials, content, instructional and assessment strategies that make students’ learning experiences highly relevant and meaningful to their lives, interests, and motivation.</li> <li>✓ The teacher uses a variety of pedagogical approaches and teaching moves that take account of, and build on, diverse language, participation and socialization patterns.</li> </ul>
<b>Moderate Level of Enactment</b>	
<p>The teacher identifies, recognizes and utilizes the experiential, cultural and linguistic resources of some students but sometimes sees diversity as a deficit or problem to be overcome. In designing activities and assessments the teacher sometimes draws on students’ prior knowledge, experiences, home and community cultures. This means that learning is sometimes not relevant and meaningful to students’ lives, interests, and motivation. The teacher cultivates relationships with some parents/caregivers and community members and includes some of them as partners in enhancing their children’s learning outcomes and opportunities.</p>	
<b>Critical Attributes</b>	<p><b>Characteristic 1. Identifies and recognizes students’ home and community culture</b></p> <ul style="list-style-type: none"> <li>✓ In some instances the teacher recognizes the diverse experiential,</li> </ul>

	<p>cultural and linguistic resources students bring to school and uses a limited number of approaches to incorporate these in the classroom and more broadly.</p> <ul style="list-style-type: none"> <li>✓ The teacher sometimes recognizes the strengths of students’ home and community cultures, but sometimes engages in deficit and stereotypical thinking.</li> <li>✓ The teacher builds positive relationships with some parents/caregivers and includes some of them as partners in enhancing their children’s learning outcomes and opportunities.</li> </ul> <p><b>Characteristic 2. Makes connections between school and home cultures by planning and implementing linguistically and culturally sensitive curriculum, instruction and assessment</b></p> <ul style="list-style-type: none"> <li>✓ The teacher sometimes embeds students’ cultural knowledge/experiences into learning experiences but is not always deliberate in the selection of materials, content, instructional and assessment strategies. This means that students’ learning experiences are sometimes not relevant or meaningful to their lives, interests, and motivation.</li> <li>✓ The teacher uses a limited number of pedagogical approaches, and teaching moves do not often take account of, and build on, diverse language, participation and socialization patterns.</li> </ul>
<b>Low Level of Enactment</b>	
<p>The teacher rarely identifies, recognizes and utilizes the diverse experiential, cultural and linguistic resources of students and usually sees diversity as a deficit or problem to be overcome. In designing activities and assessments the teacher only occasionally draws on students’ diverse prior knowledge, experiences, home and community cultures. This means that learning is often not relevant or meaningful to many students’ lives, interests, and motivation. The teacher cultivates relationships with some parents/caregivers and community members, often those who are culturally and linguistically similar to the teacher, and includes them as partners in enhancing their children’s learning outcomes and opportunities, but often excludes other parents as learning partners.</p>	
<b>Critical Attributes</b>	<p><b>Characteristic 1. Identifies and recognizes students’ home and community culture</b></p> <ul style="list-style-type: none"> <li>✓ The teacher rarely recognizes the diverse experiential, cultural and linguistic resources students bring to school and uses a very limited number of approaches to incorporate these in the classroom and more broadly.</li> <li>✓ The teacher rarely recognizes the strengths of students’ diverse home and community cultures, and often engages in deficit and stereotypical thinking.</li> <li>✓ The teacher generally does not build positive relationships with parents/caregivers and seldom includes them as partners in enhancing their children’s learning outcomes and opportunities.</li> </ul> <p><b>Characteristic 2. Makes connections between school and home cultures by planning and implementing linguistically and culturally sensitive curriculum, instruction and assessment</b></p>

	<ul style="list-style-type: none"><li>✓ The teacher rarely embeds students' cultural knowledge/experiences into learning experiences and is not deliberate in the selection of materials, content, instructional and assessment strategies. This means that students' learning experiences are often not relevant or meaningful to their lives, interests, and motivation.</li><li>✓ The teacher uses a very limited number of pedagogical approaches, and teaching moves rarely take account of, and build on, diverse language, participation and socialization patterns.</li></ul>
--	---

Facet Three

**Facet Three: Creating learning-focused, respectful and supportive learning environments**

<b>High Level of Enactment</b>	
<p>The classroom environment is characterized by highly respectful, genuinely warm and caring interactions between the teacher and students and among students. The teacher and the students collaboratively set guidelines for constructive classroom behaviors and norms. The teacher models and facilitates effective interactional skills and uses inclusive and culturally sensitive behavior management strategies. As a result, students support, motivate, and help each other to learn, and all feel a sense of belonging to the classroom and school community. The teacher creates strong academic norms by acknowledging quality work and persistence on the part of students. The teacher designs complex tasks or multi-task environment effectively to promote students collaborative problem-solving and productive dialogue, The teacher shares power and responsibility for decision-making with students to nurture their independence as learners. The teacher arranges the classroom physical space so that it is safe and accessible for all students, and efficiently and effectively orchestrates instruction, resources, and facilities to maximize students' learning.</p>	
<b>Critical Attributes</b>	<p><b>Characteristic 1: Fosters a caring, respectful, and inclusive learning environment</b></p> <ul style="list-style-type: none"> <li>✓ The interactions between the teacher and students, and among students, are highly respectful. The teacher shows genuine warmth, caring, and encouragement to all students, values the diverse backgrounds of students, and supports students to express their views.</li> <li>✓ The teacher shares power and responsibility with students by collaboratively establishing clear expectations for constructive classroom behaviors and norms which is reflected in students' self-motivated and self-disciplined learning.</li> <li>✓ The teacher consistently models and facilitates effective interactional skills so that students support, motivate, and help each other to learn, and all feel a sense of belonging to the classroom and school community.</li> <li>✓ The teacher notices and deals with negative and/or disrespectful classroom behaviors in a timely and appropriate way.</li> </ul> <p><b>Characteristic 2: Orchestrates classroom procedures and physical space to facilitate collaborative learning</b></p> <ul style="list-style-type: none"> <li>✓ The teacher deliberately arranges the classroom physical space so that it is safe, inviting and accessible for all students.</li> <li>✓ The teacher efficiently and effectively orchestrates instruction, resources, and learning spaces to maximize students' engagement and learning.</li> </ul>
<b>Moderate Level of Enactment</b>	



The classroom environment is usually respectful and warm, reflecting caring interactions between the teacher and students and among students. The teacher and the students sometimes collaboratively set guidelines for constructive classroom behaviors and norms. At times, the teacher models and facilitates effective interactional skills and sometimes uses inclusive and culturally sensitive behavior management strategies. Students sometimes support, motivate, and help each other to learn, and some students feel a sense of belonging to the classroom and school community. The teacher sometimes acknowledges quality work and persistence on the part of students. Occasionally, the teacher shares power and responsibility for decision-making with students. The teacher arranges the classroom physical space so that it is safe and accessible for some students, and generally manages instruction, resources, and facilities to enhance most students' learning.

<p><b>Critical Attributes</b></p>	<p><b>Characteristic 1: Fosters a caring, respectful, and inclusive learning environment</b></p> <ul style="list-style-type: none"> <li>✓ The interactions between the teacher and students, and among students, are generally respectful. The teacher displays warmth, caring, and encouragement to many students, values the backgrounds of some students, and supports some students to express their views.</li> <li>✓ The teacher occasionally shares power and responsibility with students by collaboratively establishing expectations for constructive classroom behaviors and norms, which is reflected in some students' self-motivation and self-disciplined learning.</li> <li>✓ The teacher sometimes models and facilitates interactional skills that support students' motivation so that only some students feel a sense of belonging to the classroom and school community.</li> <li>✓ The teacher sometimes does not notice or deal with negative and/or disrespectful classroom behaviors in a timely and appropriate way.</li> </ul> <p><b>Characteristic 2: Orchestrates classroom procedures and physical space to facilitate collaborative learning</b></p> <ul style="list-style-type: none"> <li>✓ The teacher arranges the classroom physical space so that it is safe, inviting and accessible for some students.</li> <li>✓ The teacher organizes instruction, resources, and learning spaces that engage some students in learning.</li> </ul>
-----------------------------------	--

**Low Level of Enactment**

The classroom environment is not respectful and warm, and often does not reflect caring interactions between the teacher and students and among students. The teacher and the students rarely collaboratively set guidelines for constructive classroom behaviors and norms. The teacher infrequently models and facilitates effective interactional skills and rarely uses inclusive and culturally sensitive behavior management strategies. Students do not generally support, motivate, and help each other to learn, and few students feel a sense of belonging to the classroom and school community. The teacher rarely acknowledges quality work and persistence on the part of students. The teacher does not share power and responsibility for decision-making with students. The teacher does not arrange the classroom physical space so that it is safe and accessible for all students, and generally does not manage instruction, resources, and facilities to enhance most students' learning.

learning.	
<b>Critical Attributes</b>	<p><b>Characteristic 1: Fosters a caring, respectful, and inclusive learning environment</b></p> <ul style="list-style-type: none"> <li>✓ The interactions between the teacher and students, and among students, are not generally respectful. The teacher does not display warmth, caring, and encouragement to students, does not value the backgrounds of students, and does not support students to express their views.</li> <li>✓ The teacher rarely shares power and responsibility with students by collaboratively establishing expectations for constructive classroom behaviors and norms, which is reflected in very few students' self-motivation and self-disciplined learning.</li> <li>✓ The teacher rarely models and facilitates interactional skills that support students' motivation so that few feel a sense of belonging to the classroom and school community.</li> <li>✓ The teacher almost never notices or deals with negative and/or disrespectful classroom behaviors in a timely and appropriate way.</li> </ul> <p><b>Characteristic 2: Orchestrates classroom procedures and physical space to facilitate collaborative learning</b></p> <ul style="list-style-type: none"> <li>✓ The teacher does not arrange the classroom physical space so that it is safe, inviting and accessible for all students.</li> <li>✓ The teacher does not organize instruction, resources, and learning spaces that engage students in learning.</li> </ul>

Facet Four

**Facet Four: Using evidence to scaffold learning and improve teaching**

<b>High Level of Enactment</b>	
<p>The teacher designs classroom assessment that is fully integrated into instructional activities. The teacher proactively involves students in setting specific learning goals, and students are fully aware of the assessment criteria. The evaluative climate is positive, and all students are motivated to learn and engage in self-assessing progress. The teacher circulates and interacts with all students, and provides timely, substantive, constructive, and highly responsive feedback. The teacher uses a variety of formal and informal approaches (e.g., tests, running records, questions, observations) frequently to elicit evidence of all students' learning, and uses evidence to adjust or use alternate instructional approaches to help all students learn.</p>	
<b>Critical Attributes</b>	<p><b>Characteristic 1: Design assessment, and use evidence to improve instruction</b></p> <ul style="list-style-type: none"> <li>✓ The teacher designs the classroom assessment that is well integrated into the instructional activities.</li> <li>✓ The teacher circulates and interacts with all students and uses a variety of approaches frequently to assess students' learning.</li> <li>✓ The feedback provided by the teacher is timely, substantive, constructive, and highly responsive to students.</li> <li>✓ The teacher demonstrates flexibility and responsiveness to adjust instruction and use alternate approaches.</li> </ul> <p><b>Characteristic 2: Design assessment and use evidence to scaffold learning</b></p> <ul style="list-style-type: none"> <li>✓ The teacher proactively involves students in the process of setting specific learning goals, and students are fully aware of the assessment criteria.</li> <li>✓ The evaluative climate is positive, and all students are motivated to learn and engage in self-assessing their own progress.</li> </ul>
<b>Moderate Level of Enactment</b>	
<p>The teacher designs classroom assessment that is mostly integrated into instructional activities. Usually, the teacher decides the learning goals with some involvement of students, and students appear to be aware of the assessment criteria. The evaluative climate is generally positive in guiding and directing students' learning. Sometimes the teacher circulates, interacts with students, and provides timely, substantive, constructive, and responsive feedback to students. The teacher uses some formal and informal approaches to elicit evidence of students' learning, and makes some adjustment to help students to learn.</p>	
<b>Critical Attributes</b>	<p><b>Characteristic 1: Design assessment, and use evidence to improve instruction</b></p> <ul style="list-style-type: none"> <li>✓ The teacher designs the classroom assessment that is mostly integrated into the instructional activities.</li> <li>✓ Sometimes the teacher circulates and interacts with all students and uses some formal and informal approaches to assess students' learning.</li> </ul>

	<ul style="list-style-type: none"> <li>✓ The feedback provided by the teacher is mostly timely, substantive, constructive, and responsive to students.</li> <li>✓ The teacher demonstrates some flexibility and responsiveness by sometimes adjusting instruction and using alternate approaches.</li> </ul> <p><b>Characteristic 2: Design assessment and use evidence to scaffold learning</b></p> <ul style="list-style-type: none"> <li>✓ The teacher usually decides the learning goals, and provides the students with information about the assessment criteria.</li> <li>✓ The evaluative climate is generally positive, most students are motivated to learn, and sometimes they engage in self-assessing their own progress.</li> </ul>
<b>Low Level of Enactment</b>	
<p>The classroom assessment designed by the teacher is not connected to the instructional activities. The teacher rarely involves students in the process of setting learning goals, and students are not fully aware of the assessment criteria. The evaluative climate is negative, and does not motivate students to learn or engage in self-assessment. The teacher rarely circulates and interacts with students, often ignoring students' questions and rarely providing feedback. The teacher rarely uses formal or informal approaches to gauge students' understandings or adjust instructional approaches.</p>	
<b>Critical Attributes</b>	<p><b>Characteristic 1: Design assessment, and use evidence to improve instruction</b></p> <ul style="list-style-type: none"> <li>✓ The teacher designs classroom assessment that is not integrated into instructional activities.</li> <li>✓ The teacher rarely circulates or interacts with students and rarely uses formal or informal approaches to assess students' learning.</li> <li>✓ The feedback provided by the teacher is generally not timely, substantive, constructive, and responsive to students.</li> <li>✓ The teacher demonstrates little flexibility and responsiveness in adjusting instruction and using alternate approaches.</li> </ul> <p><b>Characteristic 2: Design assessment and use evidence to scaffold learning</b></p> <ul style="list-style-type: none"> <li>✓ The teacher almost always decides the learning goals, and rarely provides the students with information about assessment criteria.</li> <li>✓ The evaluative climate is negative, with most students not being motivated to learn or engage in self-assessing their own progress.</li> </ul>

**Facet Five: Taking an inquiry stance for further professional engagement and learning**

<b>High Level of Enactment</b>	
<p>The teacher proactively reflects on his/her own practices, knowledge, biases and assumptions, and consistently takes action to support student learning outcomes even though it might not resonate with his/her existing practice or beliefs. The teacher works with others to engage in inquiry-based research on practice. The teacher identifies priorities for student learning and selects teaching strategies informed by a variety of sources (e.g., research evidence, students’ perspectives, community’s expectations, curriculum requirements,). The teacher monitors, evaluates and responds to the impact of their actions on students’ learning, The teacher has strong commitment, resilience. The teacher has a strong sense of professional identity that includes taking responsibility for student learning and their own professional development. He/she consistently advocates for positive professional learning environments and policies that support the learning of diverse students, and considers this part of the role of a teacher</p>	
<b>Critical Attributes</b>	<p><b>Characteristic 1: Takes an inquiry stance on practice:</b></p> <ul style="list-style-type: none"> <li>✓ The teacher, in collaboration with others, continuously engages in inquiry.</li> <li>✓ In identifying priorities for students the teacher builds on the perspectives of learners, and helps students take on an active role in their own learning.</li> <li>✓ The teacher uses a variety of sources (e.g., research evidence, parents/caregivers, community members, existing curriculum standards) to identify teaching strategies.</li> <li>✓ The teacher constantly monitors and evaluates the results and impacts of his/her practice on students’ learning by posing questions, collecting evidence from multiple sources, and responding appropriately.</li> <li>✓ The teacher proactively reflects on his/her own practice, knowledge, and assumptions. He/she accepts responsibility for making a difference.</li> </ul> <p><b>Characteristic 2: Collaborates with stakeholders to build supportive learning environments and professional cultures</b></p> <ul style="list-style-type: none"> <li>✓ The teacher has strong commitment to learning through inquiry, and sense of positive professional identity.</li> <li>✓ The teacher takes responsibility for what happens in their classroom and beyond by advocating for practices and policies that support the learning of diverse students.</li> </ul>
<b>Moderate Level of Enactment</b>	
<p>The teacher sometimes reflects on his/her own practices, knowledge, and assumptions, and sometimes takes action to improve student learning outcomes. The teacher sometimes collaborates with others to engage in inquiry-based research on practice. To identify the priorities for student learning and to select teaching strategies, the teacher mostly relies on certain kinds of sources such curriculum documents and his/her own teaching experiences and occasionally draws on research evidence, perspectives of</p>	

students and community expectations. The teacher sometimes follows up on the results and impacts of teaching on students' learning and makes adjustments accordingly. The teacher generally has a positive sense of professional identity. He/she sometimes advocates for positive professional learning environments and policies that support the learning of diverse students, but considers this largely peripheral to the role of a teacher.

<b>Critical Attributes</b>	<p><b>Characteristic 1: Takes an inquiry stance on practice</b></p> <ul style="list-style-type: none"> <li>✓ The teacher sometimes works with others to engage in inquiry.</li> <li>✓ In identifying priorities for students the teacher mostly relies on certain kinds of sources such curriculum documents and his/her own teaching experiences and occasionally draws on research evidence, perspectives of students and community expectations.</li> <li>✓ The teacher sometimes monitors and evaluates the results and impacts of their practice on students' learning and makes adjustments accordingly.</li> <li>✓ The teacher sometimes reflects on his/her own practice, knowledge, and assumptions. He/she generally accepts responsibility for making a difference.</li> </ul> <p><b>Characteristic 2: Collaborates with stakeholders to build supportive learning environments and professional cultures</b></p> <ul style="list-style-type: none"> <li>✓ The teacher is somewhat committed to learning through inquiry, and generally has a positive sense of professional identity.</li> <li>✓ The teacher sometimes advocates for practices and policies that support the learning of diverse students, but considers this largely peripheral to the role of a teacher.</li> </ul>
----------------------------	---

**Low Level of Enactment**

The teacher rarely reflects on his/her own practices, knowledge, and assumptions, and tends to teach in ways that he/she is familiar with. The teacher occasionally works with other colleagues to develop his/her knowledge, skills, and practices. The teacher relies on limited sources such as curriculum documents and his/her own prior experiences, for establishing student learning priorities and selecting teaching strategies. He/she occasionally investigates the results and impacts of his/her teaching on students and makes some adjustments. The teacher has little sense of his/her professional identity and rarely takes responsibility for student learning and his/her professional development. They do not advocate for practices and policies that support the learning of diverse students, and does not consider this to be part of the role of a teacher.

<b>Critical Attributes</b>	<p><b>Characteristic 1: Takes an inquiry stance on practice</b></p> <ul style="list-style-type: none"> <li>✓ The teacher rarely works with others to engage in inquiry.</li> <li>✓ In identifying priorities for students the teacher relies on limited sources such as curriculum documents and his/her own prior experiences..</li> <li>✓ The teacher occasionally investigates the results and impacts of their practice on students' learning and makes some adjustments accordingly.</li> <li>✓ The teacher rarely reflects on his/her own practice, knowledge, and assumptions and tends to teach in ways that he/she is comfortable with. He/she rarely takes the responsibility for making a difference.</li> </ul> <p><b>Characteristic 2: Collaborates with stakeholders to build supportive</b></p>
----------------------------	--

	<p><b>learning environments and professional cultures</b></p> <ul style="list-style-type: none"><li>✓ The teacher hardly shows his/her commitment to learning through inquiry, and has little sense of his/her professional identity.</li><li>✓ The teacher does not advocate for practices and policies that support the learning of diverse students, and does not consider this to be part of the role of a teacher.</li></ul>
--	---

**Facet Six: Recognizing and Challenging Inequities**

<b>High Level of Enactment</b>	
<p>The teacher consistently recognizes the importance of improving outcomes for all students and explicitly rejects deficit thinking to explain the educational under-achievement of historically marginalized students. The teacher engages regularly in critical reflection about the impact of his/her practice. She/he puts the culture of learners front and center and shares the power of decision-making and constructing knowledge with learners. The teacher strongly believes that students are able to take responsibility for their own learning and improvement. The teacher values a variety of modes of expressing content/conceptual ideas in the classroom. He/she recognizes and actively challenges taken-for-granted school practices that disadvantage certain students and privilege others. The teacher has a strong commitment to his/her professional responsibility to work with others to improve the learning and life chances of marginalized students.</p>	
<b>Critical Attributes</b>	<p><b>Characteristic 1: Recognizes and challenges one’s own deficit thinking and ways of understanding the educational experiences of historically marginalized students</b></p> <ul style="list-style-type: none"> <li>✓ The teacher consistently recognizes the importance of improving outcomes for all students and explicitly rejects deficit thinking to explain educational under achievement of historically marginalized students. The teacher engages in critical reflection on the impact of his/her practice.</li> <li>✓ She/he puts the culture of learners front and center and shares the power of decision-making and constructing knowledge with learners. The teacher strongly believes that students are able to take responsibility for their own learning and improvement. The teacher values a variety of modes of expressing content/conceptual ideas in the classroom.</li> <li>✓ He/she recognizes and actively challenges taken-for-granted school practices that disadvantage certain students and privilege others.</li> </ul> <p><b>Characteristic 2: Takes an agentic position in his/her own practice and accepts professional commitment and responsibility</b></p> <ul style="list-style-type: none"> <li>✓ The teacher has a strong commitment to his/her professional responsibility to work with others to enhance marginalized students’ learning</li> </ul>
<b>Moderate Level of Enactment</b>	
<p>The teacher recognizes the importance of improving outcomes for all students and does not accept deficit thinking to explain the educational under-achievement of historically marginalized students. The teacher sometimes engages in critical reflection on the impact of his/her practice. The teacher attempts to put the culture of learners front and center and sometimes shares the power of decision-making and constructing knowledge with learners. The teacher generally believes that students are able to take responsibility for their own learning and improvement. The teacher values some modes of expressing content/conceptual ideas in the classroom. He/she sometimes recognizes and challenges</p>	



<p>taken-for-granted school practices that disadvantage certain students and privilege others. The teacher recognizes that it is his/her professional responsibility to bring about change in marginalized students' educational achievement.</p>	
<p><b>Critical Attributes</b></p>	<p><b>Characteristic 1: Recognizes and challenges one's own deficit thinking and ways of understanding educational experiences of historically marginalized students</b></p> <p>The teacher recognizes the importance of improving outcomes for all students and does not accept deficit thinking to explain educational under-achievement of historically marginalized students. The teacher sometimes engages in critical reflection on the impact of his/her practice.</p> <ul style="list-style-type: none"> <li>✓ The teacher attempts to put the culture of learners front and center and sometimes shares the power of decision-making and constructing knowledge with learners. The teacher generally believes that students are able to take responsibility for their own learning and improvement.</li> <li>✓ The teacher values some modes of expressing content/conceptual ideas in the classroom. He/she sometimes challenges and criticizes taken-for-granted school practices that disadvantage certain students and privilege others.</li> </ul> <p><b>Characteristic 2: Takes an agentic position in his/her own practice and accepts professional commitment and responsibility</b></p> <ul style="list-style-type: none"> <li>✓ The teacher recognizes that it is his/her professional responsibility to work with others to enhance marginalized students' learning</li> </ul>
<p><b>Low Level of Enactment</b></p>	
<p>The teacher acknowledges the importance of improving outcomes for all students but views the educational under-achievement of historically marginalized students as a problem based on deficits in learners and their families and communities. The teacher is unaware of the impact of his/her own practice and the role of his/her own cultural positioning. He/she rarely shares power for decision making with students. The teacher is uncertain that students can take responsibility for their learning. Students are usually asked to learn about information given to them, and are rarely encouraged to challenge or criticize learning materials. The teacher acknowledges that there is a professional responsibility to work with others to bring about change, but has limited strategies to achieve this.</p>	
<p><b>Critical Attributes</b></p>	<p><b>Characteristic 1: Recognizes and challenges one's own deficit thinking and ways of understanding educational experiences of historically marginalized students</b></p> <ul style="list-style-type: none"> <li>✓ The teacher acknowledges the importance of improving outcomes for all students but views educational under achievement as the result of deficits in learners, their families, and their communities. The teacher is unaware of the impact of his/her own practice and cultural positioning.</li> <li>✓ He/she rarely shares decision making with students. The teacher is uncertain that all students are able to take responsibility for their own learning.</li> <li>✓ Students are usually asked to learn information given to them, and are rarely encouraged to challenge or criticize materials or information.</li> </ul> <p><b>Characteristic 2: Takes an agentic position in his/her own practice and</b></p>

	<b>accepts professional commitment and responsibility</b>
--	---

	The teacher acknowledges that there is a professional responsibility to bring about change in marginalized students' educational achievement, but has limited strategies to achieve this.
--	---

## Appendix F: Sentence Mapping Structure

*Sentence mapping structure and three levels of scenarios: Facets one, two, and six*

I. Sentence mapping structure:

(**Person**) holds (**Facet 1, Characteristic 1**). He/she sees the home culture brought by students (**Facet 2, Characteristic 1.1**), and (**Facet 2, Characteristic 1.2**). (**Person**) (**Facet 6, Characteristic 1.3**). In the classroom, (**Person**) (**Facet 2, Characteristic 2**) and (**Face 1, Characteristic 2**). He/she is (**Facet 1, Characteristic 3**).

II. Scenarios

High-level enactment (Scenario 1)

**Maria holds high expectations for all her students and communicates challenging and meaningful learning goals to all students clearly.** She sees the home culture brought by students as **the strength and assets for their learning**, and **collaborates closely with parents/caregivers and community members as the partners of students' learning.** **Maria encourages students to take initiatives on the content, assessment, and directions of learning.** In the classroom, **Maria consistently draws students' prior knowledge and culture and purposefully designs relevant and valuable learning experiences for all students.** She is **skillful in pulling together a range of instructional approaches, techniques, strategies, and moves with clear and error-free explanation to stimulate students' learning, interests, and motivations.**

Moderate-level enactment (Scenario 2)

**Tim sometimes holds inappropriately demanding or trivial expectations for students, while mostly communicates the learning goals clearly.** He sees the home culture brought by students as **the strength but sometimes engages in stereotypical thinking**, and **collaborates closely with some parents/caregivers as the partners of students' learning.** **Tim sometimes encourages students to take initiatives on the content, assessment, and directions of learning.** In the classroom, **Tim occasionally embed students' culture into learning experiences which can be limited to motivate learning and the design of learning experiences is sometimes not deliberate and irrelevant to students.** His **content knowledge is inconsistent with some errors and**

**lack of clarity, and explanation of concepts is not always compelling to capture students' interest.**

Low-level enactment (Scenario 3)

**Kevin holds high expectation for only a few students and communicates the learning goals ambiguously. He sees the home culture brought by students as an issue and weakness for their learning, and does not include parents/caregivers and community members as the partners of students' learning. Kevin determines the content, assessment, and directions of students learning with limited student involvement. In the classroom, Kevin occasionally uses students' prior knowledge and culture and rarely designs relevant and stimulating learning experiences for all students. He utilizes a limited range of instructional approaches and techniques; the explanation of key concepts contains some errors and does not capture students' interest.**

*Sentence mapping structure and three levels of scenarios: Facets two, three, and six*

I. Sentence mapping structure:

**(Person) (Facet 2, Characteristic 1.1) and (Facet 2, Characteristic 1.2). His/her interactions with students (Facet 3, Characteristic 1.1). (Person) (Facet 6, Characteristic 1.3) and (Facet 3, Characteristic 1.2). He/she (Facet 3, Characteristic 1.3). (Person) (Facet 2, Characteristic 2) and (Facet 3, Characteristic 2).**

II. Scenarios

High-level enactment (Scenario 4)

**Andrew uses students' diverse experiences as the assets to design their learning experiences, and involves parents/caregivers in students' learning process. His interaction with all students is genuinely warm, caring, and respectful. Andrew shares the decision-making power with students and co-constructs the expectations for classroom that reflect the values and backgrounds of all students. He consistently models, facilitates, and monitors effective interactional skills so that students support and motivate each other to learn in the classroom. Moreover, consistently drawing from students' prior knowledge and culture, Andrew effectively arranges the classroom space to be inviting, safe, and accessible to all students.**

Moderate-level enactment (Scenario 5)

**Tracey mostly sees students' diverse experiences as the strength for their learning though occasionally engages in stereotypical thinking, and involves some parents in students' learning process. Overall, she cares and respects her students and the interaction with them is generally warm. Tracey sometimes shares the decision-making power with students and collaborates with some to determine the classroom expectations. Once a while, she demonstrates and promotes effective interactional skills, but doesn't consistently monitor whether and how students support and motivate each other in the classroom. Occasionally drawing from students' culture and prior knowledge, Tracey is able to arrange the classroom space to be inviting, safe, and accessible to some students.**

Low-level enactment (Scenario 6)

**Christine sees students' diverse experiences as issues for their learning and does not involve parents/caregivers in students' learning process.** Her interaction with students is aloof and she mostly disregards the backgrounds of her students. **Christine rarely shares the decision-making power with students,** nor does she encourage students to express their views. She rarely models and facilitates effective interactional skills, and doesn't monitor whether students motivate and support each other in the classroom. She rarely draws students' prior knowledge and culture, and doesn't deliberately arrange the classroom space to be inviting, safe, and accessible to all students.

*Sentence mapping structure and three levels of scenarios: Facets three, four, and six*

I. Sentence mapping structure:

**(Person) (Facet 3, Characteristic 1.1). He/she (Facet 6, Characteristic 2.2), and (Facet 4, Characteristic 2.1) (Facet 3, Characteristic 1.2). He/she (Facet 3, Characteristic 2), and (Facet 4, Characteristic 1.1). Moreover, (Person) (Facet 3, Characteristic 1.3) (Facet 4, Characteristic 1.2). As a result, (Facet 4, Characteristic 2.2)**

II. Scenarios:

High-level enactment (Scenario 7)

**Katherine genuinely cares for her students, and highly values their diverse backgrounds. She strongly believes that students are capable of taking responsibility for their learning, and involves them in designing the classroom assessment and setting the expectations for classroom behaviors. She effectively arranges the instructional procedure to be engaging to all students, and integrates a variety of classroom assessment approaches well into the instructional activities. Moreover, Katherine consistently monitors students' interactions, provides timely, substantive, and constructive feedback to students, and adjusts instructional approaches based on their needs. As a result, students support each other and are motivated to set up goals and self-assess their own learning process.**

Moderate-level enactment (Scenario 8)

**Kim generally cares for her students and respects their diverse backgrounds. She generally believes that students can take some responsibility for their learning; she mostly decides the learning goals and assessment criteria for students with some occasions where she would work with students to set the classroom expectations. She arranges the instructional procedure to be engaging and accessible to some students, and integrates some forms of formal and informal classroom assessment into instructional activities. Kim occasionally monitors students' interactions, provides timely and substantive feedback to some students, and once a while adjusts instructional approaches based on their needs. As a result, most students are motivated to learn and engaged in self-assessing their own progress.**

Low-level enactment (Scenario 9)

**Sarah does not show much care and respect to students' diverse backgrounds. She does not let student take responsibility for their learning, and usually decides the learning goals, classroom expectations and assessment criteria for students and does not make them clear to students. She doesn't purposefully arrange instructional procedure to be engaging and accessible to students, and rarely integrates classroom assessment into instructional activities. Sarah rarely monitors students' classroom interactions, mostly provides marginal and procedural feedback to students, and rarely adjusts instructional approaches based on their needs. As a result, students aren't motivated to support each other's learning or to engage in the process of self-assessing their own learning.**



*Sentence mapping structure and three levels of scenarios: Facets four, five, and six*

I. Sentence mapping structure:

(Person) has (Facet 5, Characteristic 2.1); he/she (Facet 5, Characteristic 1.1). In identifying learning priorities and teaching strategies, (Person) (Facet 5, Characteristic 1.2). He/she (Facet 4, Characteristic 2.1) and (Facet 4, Characteristic 1.1; Facet 5, Characteristic 1.3). He/she (Facet 4, Characteristic 1.2). (Person's) (Facet 4, Characteristic 2.2). (Person) (Facet 5, Characteristic 2.2; Facet 6, Characteristic 2.1).

II. Scenarios:

High-level enactment (Scenario 10)

**Juan has a strong commitment to make a difference; he proactively reflects on his practice, and is willing to take unfamiliar action to support student learning.** In identifying learning priorities and teaching strategies, **Juan draws a variety of sources such as the students' culture and research evidence.** He involves students in **designing the classroom assessment and integrates fully into instruction to understand the impacts of his practice on students' learning.** He constantly interacts with and provides constructive feedback to students, and adjusts his practice appropriately. **Juan's students are motivated to learn and make progress. Juan fully embraces his responsibility to advocate for policies and practice that support diverse students' learning.**

Moderate-level enactment (Scenario 11)

**Michael is committed to make a difference; he sometimes reflects on his own practice, and takes action to support student learning.** In identifying learning priorities and teaching strategies, **Michael relies on certain kinds of sources such as standards and his own teaching experiences with minimal inputs from students' perspectives and research evidence.** He mostly designs the classroom assessment for students and sometimes integrates into the instruction to understand the impact of his practice on students' learning. **Occasionally, he circulates among students to provide feedback and makes some adjustment on his practice accordingly. Most of Michael's students are motivated to learn and make progress. Michael generally**

**accepts his responsibility to advocate for diverse students, but does not see it as essential.**

Low-level enactment (Scenario 12)

**Dave shows little commitment to make a difference; he rarely reflects on his own practice, and hardly takes action to support student learning.** In identifying learning priorities and teaching strategies, **Dave relies on limited sources such as curriculum documents and his own teaching experiences.** He almost always determines **assessment criteria on his own** and rarely integrates assessment into instruction to **understand the impact of his practice on student learning.** He seldom circulates **among students to provide feedback or adjusts his practice according to their needs.** As a result, **Dave's students aren't particularly motivated to learn and make progress.** **Dave doesn't advocate for diverse students, nor does he consider that as the responsibility of a teacher.**

*Sentence mapping structure and three levels of scenarios: Facets one, five, and six*

I. Sentence mapping structure:

**(Person) (Facet 5, Characteristic 2.1). He/she (Facet 1, Characteristic 1). He/she (Facet 6, Characteristic 1.2). In identifying learning priorities and teaching strategies, (Person) (Facet 5, Characteristic 1.2; Facet 1, Characteristic 2). He/she (Facet 1, Characteristic 3; Facet 5, Characteristic 1.3). (Person) (Facet 5, Characteristic 2.2 & 1.1).**

IV. Scenarios:

High-level enactment (Scenario 13)

**Megan has strong commitment and positive sense of professional identity to make a difference. She sets cognitively challenging goals for all students and communicates clearly. She recognizes the multiple ways of sense making and centers students' culture in the process of learning and teaching. In identifying learning priorities and teaching strategies, Megan draws a variety of sources such as community culture and research evidence to design relevant and valuable experiences for all students. She skillfully uses a range of instructional approaches, explains clearly and vividly to stimulate students' learning, and constantly monitor the results of her practice on students' learning. Megan proactively reflects on her own assumptions, willing to take unfamiliar action to respond students' needs, and sees advocating for her students as essential.**

Moderate-level enactment (Scenario 14)

**Erin generally shows some commitment and positive sense of professional identity to make a difference. She sets high but sometimes inappropriately trivial expectations for students and mostly communicates clearly. She generally recognizes students' diverse backgrounds and multiple ways of learning. In identifying learning priorities and teaching strategies, Erin uses certain kinds of source such as curriculum document and her own teaching experiences to design learning experience which can be irrelevant for some students. She uses some range of instructional approaches though her explanation contains some errors and inconsistency, and occasionally investigates the results of her teaching on student learning. Erin**

**sometimes reflects on her own assumptions, take action to support student learning, but she doesn't see advocating for her students as essential.**

Low-level enactment (Scenario 15)

**Adrian shows little commitment and sense of professional identity to make a difference. She sets cognitively challenging goals for only a few students, and doesn't usually communicate clearly. She rarely acknowledges students' culture and multiple ways of learning. In identifying learning priorities and teaching strategies, Adrian relies solely on curriculum standards and her own teaching experiences and the design of learning experiences is not stimulating and relevant to students. She uses a limited range of instructional approaches which often contain substantive errors and hardly monitors the results of her practice on students' learning. Adrian rarely reflects on her assumptions, takes actions to support student learning, nor does she see advocating for students as her responsibility.**

## **Appendix G: Guiding Questions for Feedback Sessions**

### **1. Overall**

- How long approximately did it take you to finish the entire survey?
- Did you feel fatigue at one point of answering the survey? If so, where is it in the survey?

### **2. Instructions**

- Are the instructions on responding to the scenario-style items at the beginning clear and helpful for you, and you understand how to respond to the scenarios? If not, how can I improve them?
- Do you have any wording suggestions regarding the instructions?

### **3. Scenarios and other items**

#### **3.1 For the 15 scenario-style items capturing teachers' practice for equity:**

There are reasons behind constructing the items as scenarios but I also understand the trade-off of having lengthy items like these. Questions:

- What do you think of the length of the scenarios? (They are obviously longer than items typically seen, but is it bearable from your perspective as a teacher)
- Are there any confusing/unclear wordings in the descriptions of the scenarios? Do you have any suggestions?
- Did you find yourself focusing on specific sentences in the scenarios (e.g., the beginning or the end) when reading and trying to select your response? If so, can you briefly describe your experiences and thought process at the time?

- Does it help you to choose your response based on the 5-point scale (i.e., much lower, slightly lower, about the same, etc.) when I put “consider the scenario holistically” in the instruction?

### **3.2 Demographic information:**

- Do response options for Q2, 4, 5 make sense in terms of the NZ context? Do you have any suggestions?

## Appendix H: Comparison of the Original and the Revised Scenarios

### Facets 1, 2, and 6

#### Structure of Mapping Sentences

(Person) holds (Facet 1, Characteristic 1). He/she sees students' home culture (Facet 2, Characteristic 1.1), and (Facet 2, Characteristic 1.2). (Person) (Facet 6, Characteristic 1.3). In the classroom, he/she (Facet 2, Characteristic 2) and (Face 1, Characteristic 2). (Person) (Facet 1, Characteristic 3).

Original	Revised
<p>High-level enactment: (word counts: 78)  <b>Maria holds high expectations for all students and clearly communicates challenging and meaningful learning goals.</b> She sees students' home culture as an asset and collaborates closely with parents/caregivers as partners. Maria encourages students to take initiative regarding their learning. In the classroom, she consistently draws on students' prior knowledge and culture and purposefully designs relevant learning experiences for all. Maria skillfully uses a variety of instructional approaches to motivate students' learning. Her explanations are clear, compelling, and accurate.</p>	<p><b>Maria holds high expectations for all students and clearly communicates challenging and meaningful learning goals.</b> Maria sees students' home culture as an asset and collaborates closely with parents/caregivers. She encourages students in topics that connect to their lives. She consistently draws on students' prior knowledge and culture and purposefully designs relevant learning experiences for all. Maria skillfully uses a variety of instructional approaches to motivate students' learning. Her explanations are clear, compelling, and accurate. (word counts: 74)</p>
<p>Moderate-level enactment: (word counts: 81)  <b>Tim holds high expectations for some students and mostly communicates them clearly.</b> He generally sees students' home culture as strength and collaborates with some parents/caregivers as partners. Tim sometimes encourages students to take initiative regarding their learning. In the classroom, Tim occasionally draws on students' culture to design learning experiences that are relevant to some. He uses a limited range of approaches to explain key concepts. His explanations are not always compelling and clear to all students and sometimes contain errors.</p>	<p><b>Tim holds high expectations for some students and mostly communicates them clearly.</b> Tim generally sees students' home culture as strength and collaborates with some parents/caregivers. He sometimes lets his students choose a topic of their interests to further their learning. Tim sometimes draws on cultural examples to design learning experiences that are relevant to some students. He utilizes selected number of approaches to explain key concepts. His explanations are not always compelling and clear to all students and sometimes contain errors. (word counts: 81)</p>
<p>Low-level enactment: (word counts: 84)  <b>Kevin holds high expectations for very few students and does not communicate them</b></p>	<p><b>Kevin sets achievable goals for his students but finds it hard to communicate them.</b> He sees students'</p>

<p><b><u>clearly</u></b>. He sees students' home culture as <b><u>an obstacle</u></b> and <b><u>rarely engages with parents/caregivers as partners</u></b>. <b><u>Kevin hardly encourages students to take initiative regarding their learning</u></b>. In the classroom, <b><u>Kevin seldom draws on students' prior knowledge and culture to design learning experiences that are relevant and stimulating</u></b>. He utilizes <b><u>a very limited range of instructional approaches to explain key concepts</u></b>. <b><u>His explanations contain major errors and do not capture students' interests</u></b>.</p>	<p>home culture as <b>challenging</b> and <b>doesn't expect parents to be his partners in teaching</b>. <b>Kevin sets out lessons for his students so they know what they need to do</b>. <b>Kevin uses textbooks and self-designed learning experiences that he believes deliver the appropriate curriculum</b>. He utilizes a few teaching approaches to explain key concepts. At times he feels he does not understand the concepts he is teaching well and this affects his ability to capture students' interests. (word counts: 93)</p>
---	---



Facets 2, 3, and 6

Structure of Mapping Sentences

(Person) (Facet 2, Characteristic 1 ) and (Facet 2, Characteristic 2). His/her interactions with students (Facet 3, Characteristic 1.1). (Person) (Facet 6, Characteristic 1.3) and (Facet 3, Characteristic 1.2). He/she (Facet 3, Characteristic 1.3). (Person) (Facet 3, Characteristic 2).

V1	V3
<p>Scenarios</p> <p>High-level enactment: (word counts: 83)</p> <p>Ryan <b>collaborates with</b> parents/community members and draws on students' culture and prior knowledge as valued resources to design their learning experiences. His interaction with students is genuinely warm and caring. Ryan involves students in making decisions and setting classroom expectations that are relevant to all of them. He constantly encourages and monitors <b>effective interactional skills</b> so that all students <b>support each other to learn in the classroom</b>. Moreover, Ryan effectively arranges the classroom space to be inviting, safe, and accessible to all students.</p>	<p>Ryan <b>involves</b> parents/community members in school activities and draws on students' culture as valued resources to design their learning experiences. His interaction with students is genuinely warm and caring. Ryan involves students in making decisions and setting classroom expectations that are relevant to all of them. He constantly encourages and monitors <b>supportive interactions among students</b> so that they <b>help each other and take responsibility for each other's learning</b>. Ryan effectively arranges the classroom space to be inviting, safe, and accessible to all students. (word counts: 83)</p>
<p>Moderate -level enactment: (word counts: 78)</p> <p>Tracey <b>collaborates</b> with some parents/community members and draws on some students' culture as resources to design their learning experiences. Overall, she genuinely cares for and respects her students, though occasionally still engages in stereotypical thinking. <b>Tracey and her students sometimes collaboratively set classroom expectations</b>. She <b>occasionally encourages and monitors effective interactional skills and most students support each other to learn in the classroom</b>. Additionally, Tracey arranges the classroom space to be inviting, safe, and accessible to some students.</p>	<p>Tracey <b>cooperates</b> with some parents/community members and draws on some students' culture as examples to design their learning experiences. Overall, she genuinely cares for and respects her students, though sometimes engages in stereotypical thinking. <b>Tracey sometimes involves her students in designing a lesson or setting classroom rules</b>. She <b>often has students concentrate on their own work, and sometimes encourages collaboration among students</b>. Tracey's classroom is inviting, safe, and accessible to some. (word counts: 71)</p>
<p>Low-level enactment: (word counts: 81)</p>	<p>Christine occasionally engages with</p>

<p><b>Christine rarely engages with parents/community members and sees students' culture as an obstacle for their learning. She hardly acknowledges students' diverse backgrounds and her interaction with students is mostly aloof. Christine rarely involves students in making decisions and setting classroom expectations. She hardly facilitates effective interactional skills and seldom addresses disrespectful behaviors, thus most students aren't motivated to support each other in the classroom. Also, she seldom arranges the classroom space to be inviting, safe, and accessible to students of diverse backgrounds.</b></p>	<p><b>parents/community members but generally she sees it as unnecessary. She has a quiet, reserved manner with her students, approaching all students the same way. Christine makes most of the decisions in the classroom and sets the classroom expectations. As she believes that students should work individually, Christine does not see the need to facilitate interactive skills and usually assigns work for students to carry out on their own. Her classroom is arranged to be a quiet, non-interactive space. (word counts: 82)</b></p>
--	--

Facets 3, 4, and 6

Structure of Mapping Sentences

(Person) (Facet 3, Characteristic 1.1). He/she (Facet 6, Characteristic 1.3/2.2), and (Facet 4, Characteristic 2.1) (Facet 3, Characteristic 1.2). He/she (Facet 3, Characteristic 2), and (Facet 4, Characteristic 1.1). (Person) (Facet 4, Characteristic 1.2; Facet 4, Characteristic 2.2). He/she (Facet 3, Characteristic 1.3)

V1	V3
<p>Scenarios</p> <p>High-level enactment: (word counts: 85)</p> <p><b>Katherine genuinely cares for and respects her students. She believes in her students' capacity to take initiative regarding their learning and involves them in designing assessments and setting expectations. She effectively arranges the instructional procedure to be engaging to all students, and integrates a variety of assessment approaches well into her instruction. Katherine interacts with students to provide constructive feedback and adjusts her practice appropriately. She consistently monitors and facilitates effective classroom interactions, and her students enjoy supporting each other to learn in the classroom.</b></p>	<p><b>Katherine genuinely cares for and respects her students. She encourages students to be independent learners and to investigate and build understanding of their own, and she involves them in setting criteria and goals for their learning. She effectively constructs her teaching practice to be engaging to all students, and integrates a variety of assessment approaches well into her teaching. Katherine interacts with students to provide constructive feedback and adjusts her practice appropriately. She consistently monitors and facilitates collaborative learning among her students.</b> (word counts: 82)</p>
<p>Moderate-level enactment: (word counts: 80)</p> <p><b>Kim generally cares for and respects her students. Overall, she believes in students' capacity to take initiative regarding their learning, and occasionally involves them in designing assessments and setting classroom expectations. She arranges the instructional procedure to be engaging to some students, and the assessment is mostly integrated into her instruction. Kim sometimes circulates among students to provide feedback and adjusts her practice. She occasionally monitors and facilitates classroom interactions, and most students are motivated to help each other learn.</b></p>	<p><b>Kim generally cares for and respects her students. Overall, she believes in students' capacity to take initiative regarding their learning, and sometimes involves them in designing assessments and setting classroom expectations. Her teaching practice engages some students and the assessment is generally integrated into her teaching. Kim sometimes circulates among students to provide feedback and modify her practice. She sometimes monitors and facilitates classroom interactions among students.</b> (word counts: 67)</p>

<p>Low-level enactment: (word counts: 82)  <b>Sarah shows little care and respect for her students. She rarely shares the power of decision making with students and mostly determines assessment approaches and classroom expectations on her own.</b> Her arrangement of instructional procedure is rarely engaging, and she hardly integrates assessment into her instruction. Sarah seldom interacts or gives constructive feedback to students, and demonstrates little flexibility in modifying her practice. She rarely facilitates and monitors classroom interactions and very few students are motivated to support each other to learn.</p>	<p><b>Sarah focuses her teaching on the academic side of things rather than seeing her work as a caring role. She sees herself as the authority in her classroom and decides how assessment will be carried out and how students should behave. Her teaching practice covers the curriculum, carrying out assessment to check up of student learning. Sarah uses tried and true learning activities, letting her students complete them without her involvement. She runs a quiet classroom in which students learn individually most of the time.</b> (word counts: 85)</p>
--	--

## Facets 4, 5, and 6

### Structure of Mapping Sentences

(Person) (Facet 5, Characteristic 2; Facet 6, Characteristic 2.1). (Person) (Facet 5, Characteristic 1.2) to identify learning priorities and teaching strategies. He/she (Facet 4, Characteristic 2.1) and (Facet 4, Characteristic 1.1; Facet 5, Characteristic 1.3; Facet 4, Characteristic 1.2). (Person) (Facet 5, Characteristic 1.1; Facet 4, Characteristic 2.2).

V1	V3
<p><u>Scenarios</u> High-level enactment: (word counts: 85) <b>Juan has a sense of positive professional identity and a strong commitment to bring about change that supports student learning. Juan builds on students' perspectives and draws on a variety of sources to identify learning priorities and teaching strategies. He involves students in designing assessments and fully integrates assessment into his instruction to evaluate his practice and to give feedback to students. Through constant reflections on his practice, Juan is willing to take unfamiliar action to support students and all students are motivated to learn.</b></p>	<p><b>Juan has a strong sense of professional identity as a teacher and a strong commitment to advocating on behalf of students. Juan builds on students' perspectives and draws on a variety of sources to identify learning priorities and teaching strategies. He involves students in designing assessments and fully integrates assessment into his instruction to evaluate his practice and to give feedback to students. Juan constantly reflects on his practice and tries out new approaches to motivate and respond to students' learning needs.</b> (word counts: 82)</p>
<p>Moderate-level enactment: (word counts: 85) <b>Michael generally has a positive sense of professional identity of supporting student learning, but sees advocacy as the peripheral role of a teacher. Michael relies on standards and his own experiences with minimal input from students to identify learning priorities and teaching strategies. He usually designs assessments on his own and occasionally integrates them into instruction to evaluate his practice and to give feedback to students. Michael sometimes reflects on and adjusts his practices based on students' needs, and most students are motivated to learn.</b></p>	<p><b>Michael generally has a positive sense of professional identity as a teacher, but sees advocacy on behalf of students as the peripheral role. Michael supports student learning but relies on standards and experiences with some input from students to identify learning priorities and teaching strategies. He usually designs assessments on his own and sometimes integrates them into teaching to evaluate his practice and to give feedback to students. Michael sometimes reflects on and adjusts his practice based on students' needs to better motivate their interests.</b> (word counts: 85)</p>
<p>Low-level enactment: (word counts: 80)</p>	<p><b>Dave believes good teaching is correctly</b></p>

<p>Dave has little sense of his professional identity and hardly shows commitment to support and advocate for his students. Dave relies solely on standards or curriculum documents to identify learning priorities and teaching strategies. He determines assessments on his own and rarely integrates them into his instruction to evaluate his practice and to provide feedback to students. Dave rarely reflects on his own practice and hardly responds to students' needs; very few students are motivated to learn and make progress.</p>	<p>implementing a set of techniques to ensure that students achieve curriculum expectations for their year levels. He mainly relies on standards or curriculum documents to identify learning priorities and teaching approaches. He designs assessments on his own and generally uses them to check whether students meet the minimum academic standards. Dave looks at his students' test results, sometimes altering his practice to boost their scores. (word counts: 71)</p>
--	---

Facets 1, 5, and 6

Structure of Mapping Sentences

(Person) (Facet 5, Characteristic 2.1 & 2.2). In the classroom, (Person) (Facet 1, Characteristic 1). He/she (Facet 5, Characteristic 1.2; Facet 1, Characteristic 2) and (Facet 6, Characteristic 1.2). He/she (Facet 1, Characteristic 3; Facet 5, Characteristic 1.3). (Person) (Facet 5, Characteristic 1.1).

V1	V2
<p>Scenarios</p> <p>High-level enactment: (word counts: 86)</p> <p><b>Megan fully embraces her responsibility to support and advocate for her students. In the classroom, Megan sets cognitively challenging goals and communicates clearly. She draws upon a variety of sources such as students' culture to cultivate their conceptual understanding and encourages them to challenge information in textbooks. She deliberately uses various approaches to capture students' interests and constantly monitors the results of her practice on student learning. Megan proactively reflects on her own assumptions and is willing to take unfamiliar action to respond to students' needs.</b></p>	<p><b>Megan fully embraces her responsibility to identify and challenge classroom and school practices that promote inequities for students. Megan sets cognitively challenging goals and communicates clearly and consistently. She purposefully draws upon a variety of sources to cultivate their conceptual understanding and encourages students to challenge information in textbooks. She deliberately uses various pedagogical strategies to capture students' interests. Megan also works with others in a professional community to pose questions, reflect on her own assumptions, and proactively respond to student needs.</b> (word counts: 82)</p>
<p>Moderate-level enactment: (word counts: 81)</p> <p><b>Erin generally takes responsibility to support her students' learning. In the classroom, Erin sets and communicates high expectations to some students, though learning goals are sometimes not cognitively challenging. She relies on certain sources such as curriculum documents to design learning experiences and occasionally invites students' ideas and opinions. She uses limited approaches to capture students' interests and inconsistently monitors the results of her practice. Erin sometimes reflects on her own assumptions and takes some actions to respond to students' needs.</b></p>	<p><b>Erin generally takes responsibility for supporting her students' learning. Her lesson planning is mostly guided by curriculum documents and textbooks, and she sometimes invites students' ideas and opinions. Erin relies on her familiar repertoire of teaching approaches to capture students' interests. Erin sometimes reflects on and checks if certain approaches are more effective than others in responding to student needs.</b> (word counts: 60)</p>

<p>Low-level enactment: (word counts: 83)</p> <p>Adrian shows little commitment to support and advocate for her students. In the classroom, Adrian sets cognitively challenging goals for only a few students, and doesn't communicate clearly. She relies solely on curriculum standards or her own experiences to design learning experiences and rarely encourages students to critically examine information in textbooks. She uses very limited approaches to capture students' interests and hardly monitors the results of her practice. Adrian rarely reflects on her assumptions, nor does she appropriately respond to students' needs.</p>	<p>Adrian considers her role as primarily transferring knowledge to students. Adrian sets attainable goals for students but struggles to engage them. She follows standards and curriculum documents to design her lessons and makes sure that students retain the content well. She often uses the same teaching strategies though she is unsure whether other approaches can be more or less effective for student learning. Adrian tends to work alone and sticks with what she knows. (word counts: 74)</p>
---	---



## Appendix I: Pilot Survey

### Introduction and Informed Consent

#### Welcome and Thank You!

You are invited to participate in the pilot of an instrument measuring teachers' self-reports about their enactment of practice for equity. This instrument is particularly relevant if you have some classroom teaching experiences. Please read the information below. If you agree to participate, please check the box to indicate that you understand the procedures, agree to participate, and proceed to complete the survey.

#### About the Study

The purpose of this study is to develop an instrument that captures how teachers report on their teaching practices that promote equity and social justice. This instrument is intended to be formative – that is, it is used for figuring out to what extent student teachers (US: teacher candidates) and early career teachers (US: novice teachers) say they enact equity-centered teaching in the early years of learning to teach. The study is a part of a larger research program – Rethinking Initial Teacher Education (RITE) led by researchers at the University of Auckland in New Zealand and Boston College in the United States. Your responses on this pilot instrument will provide valuable information about how to refine the instrument.

#### Procedures

If you agree to participate, you will be asked to complete an online survey. It will take approximately 15 -20 minutes to complete. There will be a series of scenario-style items regarding teaching practice and questions that get at demographic information, your teaching experience, and teaching contexts.

At the end of the survey, you will be asked whether you would like to be entered into a lottery to win one of ten \$10 Amazon gift cards. If you indicate “yes” to the question, you will be asked to provide your contact information. If you indicate “no” to the question, you will be directed to the end of the survey and there will be no identifying information attached to your responses.

#### Potential Risks

There are no expected risks to participating in this study. There may be unknown risks.

#### Potential Benefits

By being in the study, you are helping the researcher refine the instrument, which will ultimately contribute program improvement and theory building about the conditions that support teacher candidates' equity practice.

#### Costs and Compensation

There will not be any cost to you for participating in this research, other than the investment of your time. By participating in this survey, you will enter a lottery to win one of ten \$10 Amazon gift cards even if you end the study early.

**Confidentiality**

The data will be stored on a secure server. Only the principal investigator and the research supervisor, Dr. Larry Ludlow, Professor of Education Research, Measurement and Evaluation at Boston College, will have access to the data.

**Voluntary Participation/Withdrawal**

Your participation is voluntary. You are free to skip any questions or stop taking this survey at any time. There is no penalty if you do not take part or if you decide to withdraw from the study. However, if you do participate, we encourage you to complete all the questions. If you are a student at the University of Auckland, you do not jeopardize your grades or your present or future relationships with your professors and/or University of Auckland if you withdraw from the study.

**Contacts and Questions**

The researcher conducting this study is Wen-Chia Claire Chang, doctoral candidate in the Educational Research, Measurement, and Evaluation department in the Lynch School of Education at Boston College. For questions or more information concerning this research, you may contact her at [changw@bc.edu](mailto:changw@bc.edu). This research is being supervised by Dr. Larry Ludlow. He may be contacted at [ludlow@bc.edu](mailto:ludlow@bc.edu) or +1 (617) 552-4221.

If you have any questions about your rights as a research participant, you may contact the Office for Research Protections at [irb@bc.edu](mailto:irb@bc.edu) or +1 (617) 552-4778.

**Copy of Consent Form**

Please click the link below to download a copy of the Statement of Informed Consent and print for your records.

**Statement of Consent**

By checking the box below, the research team will understand that you have given your voluntary consent to participate, are aware of the survey procedures, and understand what is being asked of you.

Yes, I consent to participate

No, I do not wish to participate (Skip logic: *If 'No' is selected, skip to the end of the survey*)

>> Continue >>

## Survey instrument for the pilot

### Section I. Enactment of Practice for Equity

Instruction: Each of the scenarios below captures some aspects of teachers' practice for equity. Equity-centered teaching recognizes and challenges social and educational inequities and promotes students' learning, broadly defined to include academic, social, emotional, civic and critical learning

To respond to the scenario-style items:

- Consider each scenario holistically, reflect on your own practice, and compare your own practice against the individual teachers' practice described in each scenario.
- Based on the 5-point scale, choose one of these:
  - *About the same* means that your practice is similar to the practice of the teacher in the specific scenario;
  - *Slightly lower* or *Much lower* means that you consider your practice to be at a lower enactment level than the practice of the person in the scenario;
  - *Slightly higher* or *Much higher* means that you consider your practice to be at a higher enactment level than the practice of the person in the scenario.

Practice Item:

Joe is a teacher who has positive relationships with the parents/caregivers of some of the students in his class. He connects with some students and generally appreciates the diverse experiences they bring to school with them. Joe sometimes involves his students in helping to design a lesson or choosing a topic of their interests. Although he usually has students concentrate on their own assignments, he sometimes encourages students to work together as groups. Joe's classroom is welcoming and comfortable to some students.

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Joe's level?

Much lower    Slightly lower    About the same    Slightly higher    Much higher

>> Continue >>

**For the remainder of these items, consider each scenario holistically, reflect on your own practice, and compare your level of enactment of practice for equity against the level of the teacher described in the scenario. Based on the 5-point scale, choose one of these:**

- *About the same*: Your practice is similar to the practice of the teacher in the specific scenario;
- *Slightly lower* or *Much lower*: You do less or much less well than the practice of the person in the scenario;
- *Slightly higher* or *Much higher*: You do better or much better than the practice of the person in the scenario.

1. Tim holds high expectations for some students in his class and mostly communicates these expectations clearly. He generally sees students' home culture as a strength and collaborates with some parents/caregivers. He sometimes lets his students choose a topic consistent with their interests to further their learning. Tim sometimes draws on cultural examples to design learning experiences that are relevant to students. He utilizes a selected number of approaches to explain key concepts. His explanations, however, are not always compelling and clear to all students and sometimes contain errors. (88) (F126M)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Tim's level?

Much lower   Slightly lower   About the same   Slightly higher   Much higher

2. Katherine genuinely cares for and respects her students. She encourages students to be independent learners and to investigate and build understandings of their own, and she involves them in setting criteria and goals for their learning. She effectively constructs her teaching practice to be engaging to all students, and integrates a variety of assessment approaches into her teaching. Katherine interacts with students to provide constructive feedback and adjusts her practice appropriately. She consistently monitors and facilitates collaborative learning among her students. (81) (F346H)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Katherine's level?

Much lower   Slightly lower   About the same   Slightly higher   Much higher

3. Adrian considers her role as a teacher primarily as transmitting knowledge to students. Adrian sets attainable goals for students but struggles to engage them. She follows standards and curriculum documents to design her lessons and makes sure that students master and retain the content. She often uses the same teaching strategies although she is unsure whether other approaches would be more or less effective for student learning. Adrian tends to work alone and sticks with what she knows. (78) (F156L)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Adrian's level?

Much lower   Slightly lower   About the same   Slightly higher   Much higher

>> Continue >>

**Consider each scenario holistically, reflect on your own practice, and compare it against the individual teachers' practice described in each scenario. Based on the 5-point scale, choose one of these:**

- *About the same*: Your practice is similar to the practice of the teacher in the specific scenario;
- *Slightly lower* or *Much lower*: You do less or much less well than the practice of the person in the scenario;
- *Slightly higher* or *Much higher*: You do better or much better than the practice of the person in the scenario.

4. Ryan involves parents in school activities and draws on students' cultures as valued resources to design their learning experiences. His interaction with students is genuinely warm and caring. Ryan involves students in making decisions and setting classroom expectations that are relevant to all of them. He constantly encourages and monitors supportive interactions among students so that they help each other and take responsibility for each other's learning. Ryan effectively arranges the classroom space to be inviting, safe, and accessible to all students. (82) (F236H)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Ryan's level?

Much lower    Slightly lower    About the same    Slightly higher    Much higher

5. Dave believes good teaching is correctly implementing a set of techniques to ensure that students achieve curriculum expectations for their grade/year levels. He mainly relies on standards or curriculum documents to identify learning priorities and teaching approaches. He designs assessments on his own and generally uses them to check whether students meet the minimum academic standards. Dave reviews his students' test results, sometimes altering his practice to boost their scores. (70) (F456L)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Dave's level?

Much lower    Slightly lower    About the same    Slightly higher    Much higher

6. Kevin sets achievable goals for his students but finds it hard to communicate them. He sees students' home culture as challenging and doesn't expect parents to be his partners in teaching. Kevin sets out lessons for his students so they know what they need to do. He uses textbooks and self-designed learning experiences that he believes deliver the appropriate curriculum. He utilizes a few different teaching approaches to explain key concepts. At times he feels he does not understand the concepts he is teaching well and this affects his ability to capture students' interests. (94) (F126L)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Kevin's level?

Much lower    Slightly lower    About the same    Slightly higher    Much higher

>> Continue >>

**Consider each scenario holistically, reflect on your own practice, and compare it against the individual teachers' practice described in each scenario. Based on the 5-point scale, choose one of these:**

- *About the same*: Your practice is similar to the practice of the teacher in the specific scenario;
- *Slightly lower* or *Much lower*: You do less or much less well than the practice of the person in the scenario;
- *Slightly higher* or *Much higher*: You do better or much better than the practice of the person in the scenario.

7. Kim generally cares for and respects her students. Overall, she believes in students' capacity to take initiative regarding their learning, and sometimes involves them in designing assessments and setting classroom expectations. Her teaching practice engages some students and assessments are generally integrated into her teaching. Kim sometimes circulates among students to provide feedback and modify her practice. Accordingly She sometimes monitors and facilitates classroom interactions among students. (67) (F346M)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Kim's level?

Much lower    Slightly lower    About the same    Slightly higher    Much higher

8. Megan fully embraces her responsibility to identify and challenge classroom and school practices that promote inequities for students. Megan sets cognitively challenging goals and communicates to her students clearly and consistently. She purposefully draws upon a variety of sources to cultivate their conceptual understanding and encourages students to challenge information in textbooks. She deliberately uses various pedagogical strategies to capture students' interests. Megan also works with others in a professional community to pose questions, reflect on her own assumptions, and proactively respond to student needs. (85) (F156H)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Megan's level?

Much lower    Slightly lower    About the same    Slightly higher    Much higher

9. Christine occasionally engages with parents but generally she sees this as unnecessary. She has a quiet, reserved manner with her students, approaching all students the same way. Christine makes most of the decisions in the classroom and sets classroom expectations. Because she believes that students should work individually, Christine does not see the need to facilitate interactive skills and usually assigns work for students to carry out on their own. Her classroom is arranged to be a generally quiet, non-interactive space. (81) (F236L)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Christine's level?

Much lower   Slightly lower   About the same   Slightly higher   Much higher

>> Continue >>



**Consider each scenario holistically, reflect on your own practice, and compare it against the individual teachers' practice described in each scenario. Based on the 5-point scale, choose one of these:**

- *About the same*: Your practice is similar to the practice of the teacher in the specific scenario;
- *Slightly lower* or *Much lower*: You do less or much less well than the practice of the person in the scenario;
- *Slightly higher* or *Much higher*: You do better or much better than the practice of the person in the scenario.

10. Michael generally has a positive sense of professional identity as a teacher, but sees advocacy on behalf of students as a peripheral role. Michael supports student learning but relies on standards and personal experiences with some input from students to identify learning priorities and teaching strategies. He usually designs assessments on his own and sometimes integrates them into his teaching to evaluate his own practice and to give feedback to students. Michael sometimes reflects on and adjusts his practice based on students' needs to better motivate their interests. (88) (F456M)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Michael's level?

Much lower    Slightly lower    About the same    Slightly higher    Much higher

11. Maria holds high expectations for all students and clearly communicates challenging and meaningful learning goals. Maria sees students' home cultures as assets and collaborates closely with parents/caregivers. She encourages students to explore topics that connect to their lives. She consistently draws on students' prior knowledge and cultures and purposefully designs relevant learning experiences for all. Maria skillfully uses a variety of instructional approaches to motivate students' learning. Her explanations are clear, compelling, and accurate. (74) (F126H)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Maria's level?

Much lower    Slightly lower    About the same    Slightly higher    Much higher

12. Tracey cooperates with some parents/community members and draws on some students' culture as examples to design their learning experiences. Overall, she genuinely cares for and respects her students, though sometimes engages in stereotypical thinking. Tracey sometimes involves her students in designing a lesson or setting classroom rules. Although she often has students concentrate on their own work, she sometimes encourages collaboration among students. Tracey's classroom is inviting and safe for some students. (72) (F236M)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Tracey's level?

Much lower    Slightly lower    About the same    Slightly higher    Much higher

>> Continue >>

**Consider each scenario holistically, reflect on your own practice, and compare it against the individual teachers' practice described in each scenario. Based on the 5-point scale, choose one of these:**

- *About the same*: Your practice is similar to the practice of the teacher in the specific scenario;
- *Slightly lower* or *Much lower*: You do less or much less well than the practice of the person in the scenario;
- *Slightly higher* or *Much higher*: You do better or much better than the practice of the person in the scenario.

13. Juan has a strong sense of professional identity as a teacher and a strong commitment to advocating on behalf of students. Juan builds on students' perspectives and draws on a variety of sources to identify learning priorities and teaching strategies. He involves students in designing assessments and fully integrates assessment into his instruction to evaluate his own practice and to give feedback to students. Juan continuously reflects on his practice and tries out new approaches to motivate and respond to students' learning needs. (83) (F456H)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Juan's level?

Much lower    Slightly lower    About the same    Slightly higher    Much higher

14. Sarah focuses her teaching on the academic side of things rather than seeing the focus of her job as caring about student. She sees herself as the authority in her classroom and decides how assessment will be carried out and how students should behave. Her teaching practice covers the curriculum, and she carries out assessments to check up of student learning. Sarah uses tried and true learning activities, expecting her students complete them independently. She runs a quiet classroom in which students learn individually most of the time. (88) (F346L)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Sarah's level?

Much lower    Slightly lower    About the same    Slightly higher    Much higher

15. Erin generally takes responsibility for supporting her students' learning. Her lesson planning is mostly guided by curriculum documents and textbooks, and she sometimes invites students' ideas and opinions. Erin relies on a familiar repertoire of teaching approaches to capture students' interests. Erin sometimes reflects on and checks to see whether certain approaches are more effective than others in responding to student needs. (62) (F156M)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Erin's level?

Much lower   Slightly lower   About the same   Slightly higher   Much higher

>> Continue >>

**Section II. Social Desirability Scale**

Listed below are statements concerning personal attitudes and traits. Read each item and select "T" if the statement is true for you, or select "F" if the statement is false for you.

	True	False
1. I'm always willing to admit it when I make a mistake. (T)		
2. I always try to practice what I preach. (T)		
3. I never resent being asked to return a favor. (T)		
4. I have never been irked when people expressed ideas very different from my own. (T)		
5. I have never deliberately said something that hurt someone's feelings. (T)		
6. I like to gossip at times. (F)		
7. There have been occasions when I took advantage of someone. (F)		
8. I sometimes try to get even rather than forgive and forget. (F)		
9. At times I have really insisted on having things my own way. (F)		
10. There have been occasions when I felt like smashing things. (F)		
11. I never hesitate to go out of my way to help someone in trouble. (T)		
12. I have never intensely disliked anyone. (T)		
13. When I don't know something I don't at all mind admitting it. (T)		
14. I am always courteous, even to people who are disagreeable. (T)		
15. I would never think of letting someone else be punished for my wrong doings. (T)		
16. I sometimes feel resentful when I don't get my way. (F)		
17. There have been times when I felt like rebelling against people in authority even though I knew they were right. (F)		
18. I can remember "playing sick" to get out of something. (F)		
19. There have been times when I was quite jealous of the good fortune of others. (F)		
20. I am sometimes irritated by people who ask favors of me. (F)		

>> Continue >>

### Section III. Background Information

1. What is your gender identity?

- Male
- Female
- Others

2. What is your race/ethnicity?

- European
- Māori
- Asian
- Pasifica
- Other, please specify \_\_\_\_\_

3. What is the overall number of years you have been teaching?

- less than 1 year
- more than 1 and less than 3 years
- more than 3 and less than 5 years
- more than 5 and less than 10 years
- more than 10 years

4. What year level do you primarily teach now?

- Primary school
- Intermediate school
- Secondary school
- Other, please specify \_\_\_\_\_

5. What is the subject content area that you primarily teach now?

- English Language Arts/Literacy
- History and Social Science
- Mathematics
- Science and Technology/Engineering
- Others, please specify \_\_\_\_\_

>> Continue >>

## You have completed the survey!

Would you like to enter a lottery to win one of ten \$10 Amazon gift cards?

Yes

No (Skip logic: *If 'No' is selected, skip to the end of the block*)

Please provide your contact information (Display logic: *If 'Yes' is selected for the above question*)

Last/Family name \_\_\_\_\_ (Display logic: *If 'Yes' is selected for the above question*)

First/Given name \_\_\_\_\_ (Display logic: *If 'Yes' is selected for the above question*)

Email address \_\_\_\_\_ (Display logic: *If 'Yes' is selected for the above question*) (Request response if this text box is empty)

>> Continue >>

Thank you! You will receive an email in a couple of weeks notifying you if you have won a \$10 Amazon gift card. (Display logic: if email address is not empty)

>> Continue >>

**Thank you very much** for taking your time to participate in this survey. If you have any questions and concerns regarding the survey and/or the research, please contact the researcher, Wen-Chia Claire Chang at [changw@bc.edu](mailto:changw@bc.edu).

## Appendix J: Final Survey

### Introduction and Informed Consent

#### Welcome and Thank You!

You are invited to participate in a survey measuring teachers' self-reports about their enactment of practice for equity. This instrument is particularly relevant if you have some classroom teaching experiences. Please read the information below. If you agree to participate, please check the box to indicate that you understand the procedures, agree to participate, and proceed to complete the survey.

#### About the Study

The purpose of this study is to develop an instrument that captures how teachers report on their teaching practices that promote equity and social justice. This instrument is intended to be formative – that is, it is used for figuring out to what extent teacher candidates and novice teachers say they enact equity-centered teaching in the early years of learning to teach. The study is a part of a larger research program – Rethinking Initial Teacher Education (RITE) led by researchers at the University of Auckland in New Zealand and Boston College in the United States. Your responses on this survey will provide valuable information about how to refine the instrument.

#### Procedures

If you agree to participate, you will be asked to complete an online survey. It will take approximately 15 -20 minutes to complete. There will be a series of scenario-style items regarding teaching practice and questions that get at demographic information, your teaching experience, and teaching contexts.

At the end of the survey, you will be asked whether you would like to be entered into a lottery to win one of fifteen \$10 Amazon gift cards. If you indicate “yes” to the question, you will be asked to provide your contact information. If you indicate “no” to the question, you will be directed to the end of the survey and there will be no identifying information attached to your responses.

#### Potential Risks

There are no expected risks to participating in this study. There may be unknown risks.

#### Potential Benefits

By being in the study, you are helping the researcher refine the instrument, which will ultimately contribute program improvement and theory building about the conditions that support teacher candidates' equity practice.

#### Costs and Compensation

There will not be any cost to you for participating in this research, other than the investment of your time. By participating in this survey, you will enter a lottery to win one of fifteen \$10 Amazon gift cards even if you end the study early.



**Confidentiality**

The data will be stored on a secure server. Only the principal investigator and the research supervisor, Dr. Larry Ludlow, Professor of Education Research, Measurement and Evaluation at Boston College, will have access to the data.

**Voluntary Participation/Withdrawal**

Your participation is voluntary. You are free to skip any questions or stop taking this survey at any time. There is no penalty if you do not take part or if you decide to withdraw from the study. However, if you do participate, we encourage you to complete all the questions. If you are a student at Boston College or other university-based teacher education programs, you do not jeopardize your grades or your present or future relationships with your professors and/or the institutions if you withdraw from the study.

**Contacts and Questions**

The researcher conducting this study is Wen-Chia Claire Chang, doctoral candidate in the Educational Research, Measurement, and Evaluation department in the Lynch School of Education at Boston College. For questions or more information concerning this research, you may contact her at [changw@bc.edu](mailto:changw@bc.edu). This research is being supervised by Dr. Larry Ludlow. He may be contacted at [ludlow@bc.edu](mailto:ludlow@bc.edu) or +1 (617) 552-4221.

If you have any questions about your rights as a research participant, you may contact the Office for Research Protections at [irb@bc.edu](mailto:irb@bc.edu) or +1 (617) 552-4778.

**Copy of Consent Form**

Please click the link below to download a copy of the Statement of Informed Consent and print for your records.

[Statement of informed consent link](#)

**Statement of Consent**

By checking the box below, the research team will understand that you have given your voluntary consent to participate, are aware of the survey procedures, and understand what is being asked of you.

Yes, I consent to participate

No, I do not wish to participate (Skip logic: *If 'No' is selected, skip to the end of the survey*)

>> Continue >>

## Survey instrument

### Section I. Enactment of Practice for Equity

Instruction: Each of the scenarios below captures some aspects of teachers' practice for equity. Equity-centered teaching recognizes and challenges social and educational inequities and promotes students' learning, broadly defined to include academic, social, emotional, civic and critical learning

To respond to the scenario-style items:

- Consider each scenario holistically, reflect on your own practice, and compare your own practice against the individual teachers' practice described in each scenario.
- Based on the 5-point scale, choose one of these:
  - *About the same* means that your practice is similar to the practice of the teacher in the specific scenario;
  - *Slightly lower* or *Much lower* means that you consider your practice to be at a lower enactment level than the practice of the person in the scenario;
  - *Slightly higher* or *Much higher* means that you consider your practice to be at a higher enactment level than the practice of the person in the scenario.

Practice Item:

Joe is a teacher who has positive relationships with the parents/caregivers of some of the students in his class. He connects with some students and generally appreciates the diverse experiences they bring to school with them. Joe sometimes involves his students in helping to design a lesson or choosing a topic of their interests. Although he usually has students concentrate on their own assignments, he sometimes encourages students to work together as groups. Joe's classroom is welcoming and comfortable to some students.

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Joe's level?

Much lower    Slightly lower    About the same    Slightly higher    Much higher

>> Continue >>

A new page showing the message below will appear:

**You have just indicated that your practice for equity in the classroom is**

***“Response option selected by the participant”***

**in comparison to Joe’s level.**

**For the remainder of these items, consider each scenario holistically, reflect on your own practice, and compare your level of enactment of practice for equity against the level of the teacher described in the scenario. Based on the 5-point scale, choose one of these:**

- *About the same*: Your practice is similar to the practice of the teacher in the specific scenario;
- *Slightly lower* or *Much lower*: You do less or much less well than the practice of the person in the scenario;
- *Slightly higher* or *Much higher*: You do better or much better than the practice of the person in the scenario.

1. Tim holds high expectations for some students in his class and mostly communicates these expectations clearly. He generally sees students' home culture as a strength and collaborates with some parents/caregivers. He sometimes lets his students choose a topic consistent with their interests to further their learning. Tim sometimes draws on cultural examples to design learning experiences that are relevant to students. He utilizes a selected number of approaches to explain key concepts. His explanations are clear and interesting to all students. (81) (F126M)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Tim's level?

Much lower    Slightly lower    About the same    Slightly higher    Much higher

2. Katherine cares for and respects her students. She encourages students to be independent learners and to investigate and build understandings of their own, and she involves them in setting criteria and goals for their learning. She constructs her teaching practice to be engaging to all students, and integrates a variety of assessment approaches into her teaching. Katherine interacts with students to provide constructive feedback and adjusts her practice appropriately. She monitors and facilitates collaborative learning among her students. (79) (F346H)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Katherine's level?

Much lower    Slightly lower    About the same    Slightly higher    Much higher

3. Adrian considers her role as a teacher primarily as transmitting knowledge to students. Adrian sets attainable goals for students but struggles to engage them. She adheres to standards and curriculum documents to design her lessons and makes sure that students memorize the content. She often uses the same teaching strategies although she is unsure whether other approaches would be more or less effective for student learning. Adrian tends to work alone and sticks with what she knows. (77) (F156L)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Adrian's level?

Much lower   Slightly lower   About the same   Slightly higher   Much higher

>> Continue >>

**Consider each scenario holistically, reflect on your own practice, and compare it against the individual teachers' practice described in each scenario. Based on the 5-point scale, choose one of these:**

- *About the same*: Your practice is similar to the practice of the teacher in the specific scenario;
- *Slightly lower* or *Much lower*: You do less or much less well than the practice of the person in the scenario;
- *Slightly higher* or *Much higher*: You do better or much better than the practice of the person in the scenario.

4. Ryan involves parents in school activities and draws on students' cultures as valued resources to design their learning experiences. His interaction with students is genuinely warm and caring. Ryan involves students in making decisions and setting classroom expectations that are relevant to all of them. He constantly encourages and monitors supportive interactions among students so that they help each other and take responsibility for each other's learning. Ryan effectively arranges the classroom space to be inviting, safe, and accessible to all students. (82) (F236H)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Ryan's level?

Much lower    Slightly lower    About the same    Slightly higher    Much higher

5. Dave believes good teaching is executing a set of techniques to ensure that students attain curriculum expectations for their grade/year levels. He solely relies on standards or curriculum documents to identify learning priorities and teaching approaches. He designs assessments on his own and generally uses them to check whether students meet the minimum academic standards. Dave reviews his students' test results, sometimes altering his practice to boost their scores. (69) (F456L)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Dave's level?

Much lower    Slightly lower    About the same    Slightly higher    Much higher

6. Kevin sets achievable goals for his students but finds it hard to communicate them. He sees students' home culture as challenging and doesn't expect parents to be his partners in teaching. Kevin sets out lessons for his students so they know what they need to do. He uses textbooks and self-designed learning experiences that he believes deliver the appropriate curriculum. He utilizes a few different teaching approaches to explain key concepts. At times he feels he does not understand the concepts he is teaching well and this affects his ability to capture students' interests. (94) (F126L)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Kevin's level?

Much lower   Slightly lower   About the same   Slightly higher   Much higher

>> Continue >>

**Consider each scenario holistically, reflect on your own practice, and compare it against the individual teachers' practice described in each scenario. Based on the 5-point scale, choose one of these:**

- *About the same*: Your practice is similar to the practice of the teacher in the specific scenario;
- *Slightly lower* or *Much lower*: You do less or much less well than the practice of the person in the scenario;
- *Slightly higher* or *Much higher*: You do better or much better than the practice of the person in the scenario.

7. Kim generally cares for and respects her students. Overall, she believes in students' capacity to take initiative regarding their learning, and sometimes involves them in designing assessments and setting classroom expectations. Her teaching practice engages some students and assessments are generally integrated into her teaching. Kim sometimes circulates among students to provide feedback and modify her practice. Accordingly She sometimes monitors and facilitates classroom interactions among students. (67) (F346M)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Kim's level?

Much lower    Slightly lower    About the same    Slightly higher    Much higher

8. Megan fully embraces her responsibility to identify and challenge classroom and school practices that promote inequities for students. Megan sets cognitively challenging goals and communicates to her students clearly and consistently. She purposefully draws upon a variety of sources to cultivate their conceptual understanding and encourages students to challenge information in textbooks. She deliberately uses various pedagogical strategies to capture students' interests. Megan also works with others in a professional community to pose questions, reflect on her own assumptions, and proactively respond to student needs. (85) (F156H)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Megan's level?

Much lower    Slightly lower    About the same    Slightly higher    Much higher

9. Christine occasionally engages with parents but generally she sees this as unnecessary. She has a quiet, reserved manner with her students, approaching all students the same way. Christine, rather than involving the students, makes most of the decisions in the classroom. Because she believes that students should work individually, Christine does not see the need to facilitate interactive skills and usually assigns work for students to carry out on their own. Her classroom is arranged to be a generally quiet, non-interactive space. (82) (F236L)



Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Christine's level?

Much lower   Slightly lower   About the same   Slightly higher   Much higher

>> Continue >>

**Consider each scenario holistically, reflect on your own practice, and compare it against the individual teachers' practice described in each scenario. Based on the 5-point scale, choose one of these:**

- *About the same*: Your practice is similar to the practice of the teacher in the specific scenario;
- *Slightly lower* or *Much lower*: You do less or much less well than the practice of the person in the scenario;
- *Slightly higher* or *Much higher*: You do better or much better than the practice of the person in the scenario.

10. Michael generally has a positive sense of professional identity as a teacher, but sees advocacy on behalf of students as a peripheral role. Michael supports student learning but relies on standards and personal experiences with some input from students to identify learning priorities and teaching strategies. He usually designs assessments on his own and sometimes integrates them into his teaching to evaluate his own practice and to give feedback to students. Michael sometimes reflects on and adjusts his practice based on students' needs to better motivate their interests. (88) (F456M)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Michael's level?

Much lower    Slightly lower    About the same    Slightly higher    Much higher

11. Maria holds high expectations for all students and clearly communicates challenging and meaningful learning goals. Maria sees students' home cultures as assets and collaborates closely with parents/caregivers. She encourages students to explore topics that connect to their lives. She consistently draws on students' prior knowledge and cultures and purposefully designs relevant learning experiences for all. Maria skillfully uses a variety of instructional approaches to motivate students' learning. Her explanations are clear, compelling, and accurate. (74) (F126H)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Maria's level?

Much lower    Slightly lower    About the same    Slightly higher    Much higher

12. Tracey cooperates with some parents/community members and draws on some students' culture as examples to design their learning experiences. Overall, she genuinely cares for and respects her students, though sometimes engages in stereotypical thinking. Tracey sometimes involves her students in designing a lesson or setting classroom rules. Although she often has students concentrate on their own work, she sometimes encourages collaboration among students. Tracey's classroom is inviting and safe for some students. (72) (F236M)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Tracey's level?

Much lower   Slightly lower   About the same   Slightly higher   Much higher

>> Continue >>

**Consider each scenario holistically, reflect on your own practice, and compare it against the individual teachers' practice described in each scenario. Based on the 5-point scale, choose one of these:**

- *About the same*: Your practice is similar to the practice of the teacher in the specific scenario;
- *Slightly lower* or *Much lower*: You do less or much less well than the practice of the person in the scenario;
- *Slightly higher* or *Much higher*: You do better or much better than the practice of the person in the scenario.

13. Juan is deeply committed to supporting the learning and life of diverse students, advocating on behalf of them, and contributing to the profession. Juan builds on students' perspectives and draws on a variety of sources to identify learning priorities and teaching strategies. He involves students in designing assessments and fully integrates assessment into his instruction to provide constructive and timely feedback to students. Juan takes charge of his professional learning through continuous reflection on his practice and experimenting with new approaches to motivate and respond to students' learning needs. (88) (F456H)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Juan's level?

Much lower    Slightly lower    About the same    Slightly higher    Much higher

14. Sarah focuses her teaching on the academic side of things rather than seeing the focus of her job as caring about student. She sees herself as the authority in her classroom and decides how assessment will be carried out and how students should behave. Her teaching practice covers the curriculum, and she carries out assessments to check up of student learning. Sarah uses tried and true learning activities, expecting her students complete them independently. She runs a quiet classroom in which students learn individually most of the time. (88) (F346L)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Sarah's level?

Much lower    Slightly lower    About the same    Slightly higher    Much higher

15. Erin generally takes responsibility for supporting her students' learning. Her lesson planning is mostly guided by curriculum documents and textbooks, and she sometimes invites students' ideas and opinions. Erin relies on a familiar repertoire of teaching approaches to capture students' interests. Erin sometimes reflects on and checks to see whether certain approaches are more effective than others in responding to student needs. (62) (F156M)

Considering this scenario holistically and reflecting on your own practice, how would you describe your level of enactment of practice for equity in comparison to Erin's level?

Much lower   Slightly lower   About the same   Slightly higher   Much higher

>> Continue >>

## Section II. Social Desirability Scale

Listed below are statements concerning personal attitudes and traits. Read each item and select "T" if the statement is true for you, or select "F" if the statement is false for you. (Value: T = 1, F = 0)

	True	False
1. I'm always willing to admit it when I make a mistake. (T)		
2. I always try to practice what I preach. (T)		
3. I never resent being asked to return a favor. (T)		
4. I have never been irked when people expressed ideas very different from my own. (T)		
5. I have never deliberately said something that hurt someone's feelings. (T)		
6. I like to gossip at times. (F)		
7. There have been occasions when I took advantage of someone. (F)		
8. I sometimes try to get even rather than forgive and forget. (F)		
9. At times I have really insisted on having things my own way. (F)		
10. There have been occasions when I felt like smashing things. (F)		
11. I never hesitate to go out of my way to help someone in trouble. (T)		
12. I have never intensely disliked anyone. (T)		
13. When I don't know something I don't at all mind admitting it. (T)		
14. I am always courteous, even to people who are disagreeable. (T)		
15. I would never think of letting someone else be punished for my wrong doings. (T)		
16. I sometimes feel resentful when I don't get my way. (F)		
17. There have been times when I felt like rebelling against people in authority even though I knew they were right. (F)		
18. I can remember "playing sick" to get out of something. (F)		
19. There have been times when I was quite jealous of the good fortune of others. (F)		
20. I am sometimes irritated by people who ask favors of me. (F)		

>> Continue >>

### Section III. Background Information

1. What is your gender identity?

- Male
- Female
- Others

2. What is your race/ethnicity?

- Asian
- Black or African American
- Hispanic, Latino, or Spanish Origin
- White
- American Indian or Alaska Native
- Native Hawaiian or Other Pacific Islander
- Middle Eastern or North African
- Other origin, race, or ethnicity, please specify \_\_\_\_\_

3. What is the overall number of years you have been teaching?

- less than 1 year
- more than 1 and less than 3 years
- more than 3 and less than 5 years
- more than 5 and less than 10 years
- more than 10 years

4. What year level do you primarily teach now?

- Elementary school
- Middle school
- Secondary school
- Other, please specify \_\_\_\_\_

5. What is the subject content area that you primarily teach now?

- English Language Arts/Literacy
- History and Social Science
- Mathematics
- Science and Technology/Engineering
- Others, please specify \_\_\_\_\_

>> Continue >>

## **You have completed the survey!**

Would you like to enter a lottery to win one of ten \$10 Amazon gift cards?

Yes

No (Skip logic: *If 'No' is selected, skip to the end of the block*)

Please provide your contact information (Display logic: *If 'Yes' is selected for the above question*)

First/Given name \_\_\_\_\_ (Display logic: *If 'Yes' is selected for the above question*)

Email address \_\_\_\_\_ (Display logic: *If 'Yes' is selected for the above question*) (Request response if this text box is empty)

Thank you! You will receive an email in a couple of weeks notifying you if you have won a \$10 Amazon gift card. (Display logic: *if email address is not empty*)

>> Continue >>

**Thank you very much** for taking your time to participate in this survey. If you have any questions and concerns regarding the survey and/or the research, please contact the researcher, Wen-Chia Claire Chang at [changw@bc.edu](mailto:changw@bc.edu).



## Appendix K: WINSTEPS Program Scripts

### *Pilot Study*

```
&INST
TITLE = "TEES Pilot"
PERSON = Teacher ; persons are ...
ITEM = Scenario ; items are ...
ITEM1 = 5 ; column of response to first item in data record
NI = 16 ; number of items
NAME1 = 1 ; column of first character of person identifying label
NAMELEN = 4 ; length of person label
XWIDE = 1 ; number of columns per item response
CODES = "12345" ; valid codes in data file
T7OPTIONS=OEZ
TFILE = *
1.0 -4 4 -----
*
&END
PracticeM
F126M
F346H
F156L
F236H
F456L
F126L
F346M
F156H
F236L
F456M
F126H
F236M
F456H
F346L
F156M
END LABELS
R_3R333333345233 33
R_2q4434335325 23224
R_Tc3 35345535534254
R_A045353554 5433345
R_3K3525355435525355
R_2w3435245435434353
R_1O5535345435434254
R_3G3424245325324253
R_2T3235344424234243
R_yC5435325435424355
R_1m4425344335434343
R_pF5435345535435255
R_1r5442322434534455
R_2a4335324335333253
R_Wc3231345525434254
R_1O5535335525234254
R_1q3234255335424353
R_1I4435345435334254
R_214435334335433354
R_264135245335413354
R_7W3435125225432253
R_272135334325421355
R_DI3332324334433343
R_423334333334323344
R_6y4535355325423355
R_1c2434354435433354
R_2x4334344435434344
R_3N3434344324323243
R_5p4435355435434354
R_1m3435255435534354
R_9L4322244232334333
R_3H4435345425434244
```

R\_1G3243344434334353  
R\_D24245445435334344  
R\_3x5431345435433354  
R\_4N2335335325233344  
R\_OE3324243325313143  
R\_1D4435345335424255  
R\_2S453534434335443  
R\_9p3434245532333353  
R\_2v5434344424434244  
R\_234424345334324254  
R\_2V4545345335545445  
R\_2T2535345435534354  
R\_2V2334445435334244  
R\_Oq3445344435433353  
R\_3R4325345335344353  
R\_Y94525345335523254  
R\_2V3424254224423243  
R\_1D353535535533355  
N\_aa45353 5325424354  
N\_2Q54353454 5534354  
N\_2h5535354415524253  
N\_2P2334345325423254  
N\_3n4435344424424344  
N\_2t323533335333353  
N\_1O4335344325434354  
N\_T7323243332233233  
N\_3J4435245435534354  
N\_4S4325234335324354  
N\_2P4435445535434323  
N\_1I3425344324513144  
N\_3g2334344335531233  
N\_2a5435355525534354  
N\_2Z3435345334333334  
N\_3D3334334324423343  
N\_1h3325355425524254  
N\_1g4435344435435355  
N\_06553535555535355  
N\_sX3345344435334243  
N\_3N3323343444334433  
N\_Cl323433334333253  
N\_262535355435435353

*Final Study (57 cases)*

```
&INST
TITLE = "TEES Pilot"
PERSON = Teacher ; persons are ...
ITEM = Scenario ; items are ...
ITEM1 = 5 ; column of response to first item in data record
NI = 16 ; number of items
NAME1 = 1 ; column of first character of person identifying label
NAMELEN = 4 ; length of person label
XWIDE = 1 ; number of columns per item response
CODES = "12345" ; valid codes in data file
T7OPTIONS=OEZ
TFILE = *
1.0 -4 4 -----
*
&END
PracticeM
F126M
F346H
F156L
F236H
F456L
F126L
F346M
F156H
F236L
F456M
F126H
F236M
F456H
F346L
F156M
END LABELS
R_a643243533354 2143
R_eL4335345425514354
R_Br133333333333222
R_3r5435255425424353
R_Qa4335334434444354
R_I04334244334323344
R_1o3334245425423254
R_2s3334355435535354
R_yD3435255435424355
R_1C3335344335534354
R_2T3334344324333343
R_xs4325354314343153
R_3d4335324535323254
R_1L3235345335333355
R_3J3244344334433243
R_264323233243343234
R_T73245435435434325
R_uf3334345425433353
R_Bx3324354425434343
R_224335354325424154
R_2c333535335434354
R_1G12242343 4323243
R_854335345325424254
R_3F3324245235434354
R_835535345435523254
R_2r433535435535354
R_Ra4541431555455444
R_1K3224344424423214
R_eU4435355435534355
R_2c3335354434433354
R_2d4435355425533255
R_Wv4334354435435355
R_wY5435355535525355
R_Ij2345355435534354
R_743321441442434312
```

R\_293435445435534353  
R\_2d4335345435524354  
R\_2c4435355425434355  
R\_313335345335433344  
R\_1g443535535535355  
R\_1Q4435355425424255  
R\_2a4345355435324455  
R\_2z543535535434354  
R\_9v4435345435533354  
R\_1C433233434243432  
R\_2b4345355425424344  
R\_3R4335345425333354  
R\_2u3335245225333243  
R\_2Y2234244314323143  
R\_1i5435345325423354  
R\_253441322 51332312  
R\_2D3435335425434354  
R\_1I5345355435434355  
R\_033335345325434254  
R\_2w2324334324323233  
R\_Td3334344435434243  
R\_2Q343241423233312

*Final Study (52 cases)*

```
&INST
TITLE = "TEES Pilot"
PERSON = Teacher ; persons are ...
ITEM = Scenario ; items are ...
ITEM1 = 5 ; column of response to first item in data record
NI = 15 ; number of items
NAME1 = 1 ; column of first character of person identifying label
NAMELEN = 4 ; length of person label
XWIDE = 1 ; number of columns per item response
CODES = "12345" ; valid codes in data file
T7OPTIONS=OEZ
TFILE = *
1.0 -4 4 -----
*
&END
F126M
F346H
F156L
F236H
F456L
F126L
F346M
F156H
F236L
F456M
F126H
F236M
F456H
F346L
F156M
END LABELS
R_a63243533354 2143
R_eL335345425514354
R_Br33333333333222
R_3r435255425424353
R_Qa335334434444354
R_1O334244334323344
R_1o334245425423254
R_2s334355435535354
R_yD435255435424355
R_1C335344335534354
R_2T334344324333343
R_xs325354314343153
R_3d335324535323254
R_1L235345335333355
R_3J244344334433243
R_26323233243343234
R_T7245435435434325
R_uf334345425433353
R_Bx324354425434343
R_22335354325424154
R_2c335335335434354
R_1G2242343 4323243
R_85335345325424254
R_3F324245235434354
R_83535345435523254
R_2r335335435535354
R_1K224344424423214
R_eU435355435534355
R_2c335354434433354
R_2d435355425533255
R_Wv334354435435355
R_wY43535535525355
R_1j345355435534354
R_29435445435534353
R_2d335345435524354
R_2c435355425434355
```

R\_31335345335433344  
R\_1g435355535535355  
R\_1Q435355425424255  
R\_2a345355435324455  
R\_2z435355535434354  
R\_9v435345435533354  
R\_2b345355425424344  
R\_3R335345425333354  
R\_2u335245225333243  
R\_2Y234244314323143  
R\_1i435345325423354  
R\_2D435335425434354  
R\_11345355435434355  
R\_O3335345325434254  
R\_2w324334324323233  
R\_Td334344435434243

## Appendix L: Revised Teaching Equity Enactment Scenarios in “Difficulty” Order

Scenario Name (Facet/Level) & Number	Scenarios	Facet Score
F156H (#13)	Megan fully embraces her responsibility to identify and challenge classroom and school practices that promote inequities for students. Megan sets cognitively challenging goals and communicates to her students clearly and consistently. She purposefully draws upon a variety of sources to cultivate their conceptual understanding and encourages students to challenge information in textbooks. She deliberately uses various pedagogical strategies to capture students’ interests. Megan also works with others in a professional community to pose questions, reflect on her own assumptions, and proactively respond to student needs.	9
F456H (#10)	Juan is deeply committed to supporting the learning and life of diverse students, advocating on behalf of them, and contributing to the profession. Juan builds on students’ perspectives and draws on a variety of sources to identify learning priorities and teaching strategies. He involves students in designing assessments and fully integrates assessment into his instruction to provide constructive and timely feedback to students. Juan takes charge of his professional learning through continuous reflection on his practice and experimenting with new approaches to motivate and respond to students’ learning needs.	9
F126H (#1)	Maria holds high expectations for all students and clearly communicates challenging and meaningful learning goals. Maria sees students’ home cultures as assets and collaborates closely with parents/caregivers. She encourages students to explore topics that connect to their lives. She consistently draws on students’ prior knowledge and cultures and purposefully designs relevant learning experiences for all. Maria skillfully uses a variety of instructional approaches to motivate students’ learning. Her explanations are clear, compelling, and accurate.	9
F236H (#4)	Ryan involves parents in school activities and draws on students’ cultures as valued resources to design their learning experiences. His interaction with students is genuinely warm and caring. Ryan involves students in making decisions and setting classroom expectations that are relevant to all of them. He constantly encourages and monitors supportive interactions among students so that they help each other and take responsibility for each other’s learning. Ryan effectively arranges the classroom space to be inviting, safe, and accessible	9

	to all students.	
F346H (#7)	Katherine cares for and respects her students. She encourages students to be independent learners and to investigate and build understandings of their own, and she involves them in setting criteria and goals for their learning. She constructs her teaching practice to be engaging to all students, and integrates a variety of assessment approaches into her teaching. Katherine interacts with students to provide constructive feedback and adjusts her practice appropriately. She monitors and facilitates collaborative learning among her students.	8
F126M (#2)	Tim holds high expectations for some students in his class and mostly communicates these expectations clearly. He generally sees students' home culture as a strength and collaborates with some parents/caregivers. He sometimes lets his students choose a topic consistent with their interests to further their learning. Tim sometimes draws on cultural examples to design learning experiences that are relevant to students. He utilizes a selected number of approaches to explain key concepts. His explanations are clear and interesting to all students.	7
F236M (#5)	Tracey cooperates with some parents/community members and draws on some students' culture as examples to design their learning experiences. Overall, she genuinely cares for and respects her students, though sometimes engages in stereotypical thinking. Tracey sometimes involves her students in designing a lesson or setting classroom rules. Although she often has students concentrate on their own work, she sometimes encourages collaboration among students. Tracey's classroom is inviting and safe for some students.	6
F346M (#8)	Kim generally cares for and respects her students. Overall, she believes in students' capacity to take initiative regarding their learning, and sometimes involves them in designing assessments and setting classroom expectations. Her teaching practice engages some students and assessments are generally integrated into her teaching. Kim sometimes circulates among students to provide feedback and modify her practice. Accordingly She sometimes monitors and facilitates classroom interactions among students.	6
F156M (#14)	Erin generally takes responsibility for supporting her students' learning. Her lesson planning is mostly guided by curriculum documents and textbooks, and she sometimes invites students' ideas and opinions. Erin relies on a familiar repertoire of teaching approaches to capture students' interests. Erin sometimes reflects on and checks to see whether certain approaches are more effective than others in responding to student needs.	6
F456M	Michael generally has a positive sense of professional identity	6



(#11)	as a teacher, but sees advocacy on behalf of students as a peripheral role. Michael supports student learning but relies on standards and personal experiences with some input from students to identify learning priorities and teaching strategies. He usually designs assessments on his own and sometimes integrates them into his teaching to evaluate his own practice and to give feedback to students. Michael sometimes reflects on and adjusts his practice based on students' needs to better motivate their interests.	
F456L (#12)	Dave believes good teaching is executing a set of techniques to ensure that students attain curriculum expectations for their grade/year levels. He solely relies on standards or curriculum documents to identify learning priorities and teaching approaches. He designs assessments on his own and generally uses them to check whether students meet the minimum academic standards. Dave reviews his students' test results, sometimes altering his practice to boost their scores.	3
F346L (#9)	Sarah focuses her teaching on the academic side of things rather than seeing the focus of her job as caring about student. She sees herself as the authority in her classroom and decides how assessment will be carried out and how students should behave. Her teaching practice covers the curriculum, and she carries out assessments to check up of student learning. Sarah uses tried and true learning activities, expecting her students complete them independently. She runs a quiet classroom in which students learn individually most of the time.	3
F126L (#3)	Kevin sets achievable goals for his students but finds it hard to communicate them. He sees students' home culture as challenging and doesn't expect parents to be his partners in teaching. Kevin sets out lessons for his students so they know what they need to do. He uses textbooks and self-designed learning experiences that he believes deliver the appropriate curriculum. He utilizes a few different teaching approaches to explain key concepts. At times he feels he does not understand the concepts he is teaching well and this affects his ability to capture students' interests.	3
F156L (#15)	Adrian considers her role as a teacher primarily as transmitting knowledge to students. Adrian sets attainable goals for students but struggles to engage them. She adheres to standards and curriculum documents to design her lessons and makes sure that students memorize the content. She often uses the same teaching strategies although she is unsure whether other approaches would be more or less effective for student learning. Adrian tends to work alone and sticks with what she knows.	3
F236L (#6)	Christine occasionally engages with parents but generally she sees this as unnecessary. She has a quiet, reserved manner with	3

	<p>her students, approaching all students the same way. Christine, rather than involving the students, makes most of the decisions in the classroom. Because she believes that students should work individually, Christine does not see the need to facilitate interactive skills and usually assigns work for students to carry out on their own. Her classroom is arranged to be a generally quiet, non-interactive space.</p>	
--	---	--

## Appendix M: Scoring Conversing Table

TABLE OF SAMPLE NORMS (500/100) AND FREQUENCIES CORRESPONDING TO COMPLETE TEST

SCORE	MEASURE	S.E.	NORMED	S.E.	FREQUENCY	%	CUM.FREQ.	%	PERCENTILE
15	-10.49E	1.88	-534	154	0	.0	0	.0	0
16	-9.14	1.10	-424	90	0	.0	0	.0	0
17	-8.23	.85	-349	70	0	.0	0	.0	0
18	-7.60	.75	-297	62	0	.0	0	.0	0
19	-7.08	.69	-254	57	0	.0	0	.0	0
20	-6.62	.65	-217	54	0	.0	0	.0	0
21	-6.22	.62	-184	51	0	.0	0	.0	0
22	-5.84	.60	-154	49	0	.0	0	.0	0
23	-5.50	.58	-125	47	0	.0	0	.0	0
24	-5.17	.57	-98	46	0	.0	0	.0	0
25	-4.85	.56	-73	46	0	.0	0	.0	0
26	-4.55	.55	-47	45	0	.0	0	.0	0
27	-4.25	.55	-23	45	0	.0	0	.0	0
28	-3.95	.54	1	44	0	.0	0	.0	0
29	-3.67	.53	25	44	0	.0	0	.0	0
30	-3.38	.53	48	43	0	.0	0	.0	0
31	-3.11	.52	70	43	0	.0	0	.0	0
32	-2.84	.51	92	42	0	.0	0	.0	0
33	-2.58	.50	113	41	0	.0	0	.0	0
34	-2.33	.50	134	41	0	.0	0	.0	0
35	-2.09	.49	154	40	0	.0	0	.0	0
36	-1.85	.48	173	40	0	.0	0	.0	0
37	-1.62	.48	192	39	0	.0	0	.0	0
38	-1.39	.47	211	39	0	.0	0	.0	0
39	-1.17	.47	229	38	0	.0	0	.0	0
40	-.96	.46	247	38	0	.0	0	.0	0
41	-.74	.46	264	38	0	.0	0	.0	0
42	-.53	.46	282	38	1	1.9	1	1.9	1
43	-.32	.46	299	37	2	3.8	3	5.8	4
44	-.11	.45	316	37	2	3.8	5	9.6	8
45	.09	.45	332	37	1	1.9	6	11.5	11
46	.29	.45	349	37	0	.0	6	11.5	12
47	.50	.45	366	37	1	1.9	7	13.5	13
48	.70	.45	382	37	0	.0	7	13.5	13
49	.90	.45	398	37	4	7.7	11	21.2	17
50	1.10	.45	415	37	1	1.9	12	23.1	22
51	1.30	.45	432	37	0	.0	12	23.1	23
52	1.51	.45	448	37	2	3.8	14	26.9	25
53	1.72	.46	465	38	4	7.7	18	34.6	31
54	1.93	.46	483	38	2	3.8	20	38.5	37
55	2.15	.47	501	39	5	9.6	25	48.1	43
56	2.37	.48	519	39	6	11.5	31	59.6	54
57	2.60	.49	538	40	3	5.8	34	65.4	63
58	2.85	.50	558	41	3	5.8	37	71.2	68
59	3.10	.51	578	41	5	9.6	42	80.8	76
60	3.36	.51	600	42	4	7.7	46	88.5	85
61	3.63	.52	622	43	3	5.8	49	94.2	91
62	3.91	.53	645	44	1	1.9	50	96.2	95
63	4.19	.54	668	44	1	1.9	51	98.1	97
64	4.49	.55	693	45	1	1.9	52	100.0	99
65	4.80	.56	718	46	0	.0	52	100.0	100
66	5.12	.57	744	47	0	.0	52	100.0	100
67	5.45	.58	771	47	0	.0	52	100.0	100
68	5.79	.59	799	48	0	.0	52	100.0	100
69	6.15	.61	829	50	0	.0	52	100.0	100
70	6.53	.63	860	52	0	.0	52	100.0	100
71	6.96	.67	895	55	0	.0	52	100.0	100
72	7.44	.73	934	59	0	.0	52	100.0	100
73	8.03	.83	983	68	0	.0	52	100.0	100
74	8.91	1.09	1055	89	0	.0	52	100.0	100
75	10.24E	1.87	1163	153	0	.0	52	100.0	100