



University of Kentucky  
UKnowledge

---

Theses and Dissertations--Education Sciences

College of Education

---


2022

## Validation of An Early Childhood Emotional and Social Competence Instrument Using a Rasch Partial Credit Model

Nan Li

University of Kentucky, [linan.uky@gmail.com](mailto:linan.uky@gmail.com)

Author ORCID Identifier:

 <https://orcid.org/0000-0002-3932-7057>

Digital Object Identifier: <https://doi.org/10.13023/etd.2022.380>

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

### Recommended Citation

Li, Nan, "Validation of An Early Childhood Emotional and Social Competence Instrument Using a Rasch Partial Credit Model" (2022). *Theses and Dissertations--Education Sciences*. 116.  
[https://uknowledge.uky.edu/edsc\\_etds/116](https://uknowledge.uky.edu/edsc_etds/116)

This Doctoral Dissertation is brought to you for free and open access by the College of Education at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Education Sciences by an authorized administrator of UKnowledge. For more information, please contact [UKnowledge@lsv.uky.edu](mailto:UKnowledge@lsv.uky.edu).

## **STUDENT AGREEMENT:**

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

## **REVIEW, APPROVAL AND ACCEPTANCE**

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Nan Li, Student

Dr. Kelly D. Bradley, Major Professor

Dr. Jane McEldowney Jensen, Director of Graduate Studies

Validation of An Early  
Childhood Emotional and Social Competence  
Instrument Using a Rasch Partial Credit Model

---

DISSERTATION

---

A dissertation submitted in partial fulfillment of the  
requirements for the degree of Doctor of Philosophy in the  
College of Education  
at the University of Kentucky

By  
Nan Li  
Lexington, Kentucky  
Director: Dr. Kelly D. Bradley, Professor of Policy Studies and Evaluation  
Lexington, Kentucky  
2022

Copyright © Nan Li 2022  
<https://orcid.org/0000-0002-3932-7057>

## ABSTRACT OF DISSERTATION

### Validation of An Early Childhood Emotional and Social Competence Instrument Using a Rasch Partial Credit Model

The release of the Every Student Succeeds Act (ESSA) started a new chapter of education in the United States. The ESSA emphasizes the importance of preschool education; the act states that more resources will be devoted to preschool education, which lays the foundation for children's school learning process. The goal of the act is to make sure children enter kindergarten education with equal preparatory. School, family, and researchers are paying more attention to developing preschool children's social and emotional competency before entering Kindergarten. For understanding preschool children's social and emotional skill development situation, instrument is the mostly widely adopted method by researchers. The quality of the instrument adopted to measure children's' competency play an important role in getting reliable measurement results.

Therefore, the Rasch model validation method was adopted in this study to assess The Feeling Friend instrument which was designed to measure preschool children's social and emotional ability. Rasch model remedy multiple flaws of classic test theory validation method. The Rasch model can provide multiple validity and reliability diagnoses from multiple perspectives. In this study, the Rasch model validation diagnoses including reliability, validity, single item fit, DIF, and category threshold are adopted to provide a validation guide for researchers to apply the Rasch model in assessing their self-design instrument based on their research needed. The Rasch model can also provide a solution to assess instruments with a small item sample or for a small number sample, such as an evaluation of instrument quality, single item fit, and rating scale. In this study, equating method (stacking) was then adopted to assess the children's competency in social and emotional change over time.

Through use of Rasch validation model, researchers can better design and modify instruments ensure measurement quality, and the results collected from the instrument would be more reliable and trustable.

KEYWORDS: Preschool, Social and Emotional Competency, Rasch, Instrument  
Validation, Stacking Comparison

Nan Li

---

*(Name of Student)*

11/21/2022

---

Date

Validation of An Early  
Childhood Emotional and Social Competence  
Instrument Using a Rasch Partial Credit Model

By  
Nan Li

---

Dr. Kelly D. Bradley  
Director of Dissertation

---

Dr. Jane McEldowney Jensen  
Director of Graduate Studies

---

11/21/2022

Date

## DEDICATION

*To my deceased grandfather and grandmother:  
Without your unconditional love, I can't  
be the person I want to be.  
I always miss  
You!*

## ACKNOWLEDGMENTS

The following dissertation, while an individual work, benefited from the insights and direction of several people. First, my Dissertation Chair, the best advisor, Dr. Kelly Bradley, guided me to complete my learning goal at the University of Kentucky and provided me with strong support during my learning period. Your encouragement and support have helped me overcome the most difficult period in my life and created my best learning experience. My Ph. D learning period with you has been the happiest time in my lifetime.

I am also grateful to my committee member Dr. Peabody, who taught me a lot about the Rasch methodology and provided me valuable feedback in my studying. I am also thankful to Dr. John Thelin, your questions from a historical perspective enrich the context and content of my study. Additionally, I am grateful to Dr. Mark Liu, your questions always gave me a new way of thinking and gave me motivation to learn more. Your opinions about my future career are very meaningful and thoughtful. Lastly, I would like to thank Dr. Waddington for joining my committee and helping me complete my dissertation.

I also want to thank Dr. Shannon Sampson and Josh. The work experience in the evaluation center has been memorable, and I appreciate the valuable feedback from you both. The working experience prepared me for my future career path.

Lastly, I would like to thank all my friends and colleagues, including Dr. Rui Jin; Dr. Chunling Niu; Dr. Rongxiu Wu; Min Xiao; and Jing Zhang. I am so honored to learn with you. It was a long journey getting here. I am so appreciative to have you accompanying me on this journey.



## TABLE OF CONTENTS

ACKNOWLEDGMENTS .....	iii
LIST OF TABLES .....	vii
LIST OF FIGURES .....	viii
CHAPTER 1.INTRODUCTION .....	1
1.1 Early Childhood Education.....	1
1.2 Preschool Social and Emotional Education with School Readiness.....	3
1.3 Importance In Measurement of Social and Emotional Competency .....	5
1.4 Instrument Quality .....	7
1.5 Education Performance Assessment .....	9
1.6 Purpose.....	9
1.7 Research Questions.....	13
1.8 Overview of Subsequent Dissertation Chapters .....	14
CHAPTER 2.LITERATURE REVIEW .....	16
2.1 Social and Emotional Learning Background .....	16
2.1.1 Social and Emotional Definition.....	16
2.1.2 Social and Emotional Education History .....	19
2.1.3 Policy and Standard .....	22
2.1.4 social and emotional learning Accountability.....	24
2.1.5 Importance of Social and Emotional Education .....	25
2.1.6 Social And Emotional Scales.....	26
2.1.7 Social and Emotional Development Intervention Program .....	28
2.2 Rasch.....	31
2.2.1 Rasch Measurement Theory .....	31
2.2.2 Rasch model assumption.....	33
2.2.3 Rasch Partial Credit Model.....	34
2.3 Assess Results from Different Time Point.....	35
2.4 Summary .....	37
CHAPTER 3.METHODOLOGY .....	39
3.1 Research questions.....	40
3.2 Sample.....	41
3.2.1 Participants.....	41
3.3 The Feeling Friend Instrument.....	43

3.4 Instrument Construct Validity.....	44
3.4.1 Content Validity.....	44
3.4.2 Rasch Assumption Test.....	45
3.4.3 Rasch Model in Transfer Ordinal Data to Equal Interval Data .....	46
3.4.4 Rasch Partial Credit Model.....	47
3.4.5 Instrument validity and reliability.....	48
3.4.6 Individual Item Quality test .....	48
3.4.6.1 Point- Measure Correlation.....	49
3.4.6.2 Item Fit.....	49
3.4.7 Item and Person parameter estimation.....	50
3.4.8 Differential Item Functioning (DIF) .....	51
3.4.9 Category threshold.....	52
3.5 Assessment of Children's social and emotional Competency Change Over Time	53
3.6 Wright Map (Variable Map) and KeyForm .....	57
3.6.1 Wright Map (Variable Map).....	57
3.6.2 Keyform .....	59
3.7 Chapter Summary .....	59
CHAPTER 4.RESULTS.....	61
4.1 Content Validity.....	61
4.2 Model Assumption Test.....	64
4.2.1 Unidimensionality of TFF instrument .....	64
4.3 Overall Instrument Quality .....	65
4.4 Individual Item Quality Investigation.....	66
4.4.1 Point- Measure Correlation.....	66
4.4.2 Item Fit.....	67
4.5 Differential Item Functioning (DIF).....	68
4.5.1 DIF by Genders.....	68
4.5.2 DIF by Sites .....	70
4.6 Category Andrich Diagnostics.....	73
4.7 The Feeling Friends Instrument Revised .....	76
4.7.1 Misfit Item Remove Procedure.....	76
4.7.2 Revised Unidimensionality Test Result.....	76
4.7.3 Revised Item Measure and Fit .....	80
4.7.4 Wright Map.....	81
4.7.5 Keyform .....	84
4.8 Pre and Post Assessments Result Comparison .....	84
4.8.1 Stacking Result .....	85
4.8.2 T-test .....	86
4.9 Chapter Summary .....	87
CHAPTER 5.CONCULSION AND DISCUSSION .....	90

5.1	Research Question Results.....	92
5.1.1	Research Question One.....	92
5.1.2	Research Question Two.....	93
5.1.3	Research Question Three.....	95
5.1.4	Research Question Four.....	95
5.2	Discussion.....	96
5.2.1	Stacking Comparison Assessment Tool.....	96
5.3	Contribution.....	98
5.4	Limitations.....	99
5.4.1	Item Range.....	99
5.4.2	Control Group Design.....	99
	APPENDICES.....	100
	APPENDIX 1. ORIGINAL TFF INSTRUMENT.....	100
	APPENDIX 2. REVISED TFF INSTRUMENT.....	102
	REFERENCES.....	103
	VITA.....	118

## LIST OF TABLES

Table 3.1: Sample Distribution.....	43
Table 4.1: Item Group based on CASEL and NC social and emotional preschool learning standard.....	63
Table 4.2: Standardized Residual Variance.....	65
Table 4.3: Reliability and Separation of Instrument.....	66
Table 4.4: Item measure and Fit assessment.....	67
Table 4.5: Pairwise DIF Analysis of Gender:.....	69
Table 4.6: Pairwise DIF Analysis of Sites:.....	71
Table 4.7: Disorder Andrich Threshold.....	74
Table 4.8: Iterative Procedures to Access of Item Fit.....	76
Table 4.9: Revised Unidimensionality Result.....	77
Table 4.10: First Contrast Loading and Cluster Group.....	78
Table 4.11: Correlation Coefficient of Item Clusters within First Contrast.....	80
Table 4.12: Revised Item Measure and Fit.....	81
Table 4.13: pre- and post- t-test result.....	87

## LIST OF FIGURES

Figure 2.1: Interactive CASEL Wheel.....	18
Figure 2.2: Comer SDP Model .....	21
Figure 3.1: Stacking Data Structure.....	57
Figure 4.1: DIF Measure Plot of pairwise comparison group by Gender.....	70
Figure 4.2: DIF Measure Plot of pairwise comparison group by Sites.....	73
Figure 4.3: measure relative to item difficulty.....	75
Figure 4.4: Revised First Contrast Standardized Residual Plot.....	79
Figure 4.5: Wright Map .....	83
Figure 4.6: Keyform.....	84
Figure 4.7: Stacking Comparison Plot.....	86

## CHAPTER 1. INTRODUCTION

### 1.1 Early Childhood Education

The release of the Every Student Succeeds Act (ESSA) started a new chapter of education in the United States. The ESSA emphasizes the importance of preschool education; the act states that more resources will be devoted to preschool education, which lays the foundation for children's school learning process, to make sure children enter kindergarten education with equal preparatory. The Institute of Education Science (IES) states, "The goal of early childhood education is to improve academic and social-behavioral outcomes for children from birth through third grade" (D. E. Jones et al., 2015).

As early as 1960, education for young children was considered a family responsibility. The fast development and important years period of childhood are between 0-6 years old, and so early childhood education will provide a positive influence on children's whole life well-being (Cannon et al., 2017). In this period, the objectives of education are to help children prepare for primary education and build good learning behavior, and to develop positive habits. Generally, children can optionally participate in preschool education training when they are aged approximately 3 to 5 years old in the United States. Preschool education is the first stage of education in children's life, just before kindergarten education. It provides essential training in cognitive functioning and socialization by professionally trained adults.

Head Start is a federally funded education program which targets households with economic difficulties and children in need to aid childhood development. Head Start defining readiness as "children possessing the skill, knowledge, and attitudes necessary for

success in school and for later learning and life. Physical, cognitive, social, and emotional development are essential ingredients of school readiness” (School Readiness | ECLKC, n.d.). Research in this area continues to advance, and a coherent understanding of the links between children’s readiness and their future academic achievement is increasing: early education programs improve children’s school readiness by giving them the academic skills for kindergarten education. Magnuson et al. (2004) suggests that children who participated in preschool education tend to perform better in kindergarten education, and these children had a lower Kindergarten repeat rate. A well-designed preschool program can help children develop skills in language, socialization, and cognitive ability.

Preschool education can benefit children in multiple aspects. Children with preschool education have more stable home lives, and are more responsible as citizens (*Preschool Education | Encyclopedia.Com, 2018*). Large-scale longitudinal studies provide accumulating evidence to support that preschool education could benefit children with a long-lasting effect in cognitive ability, school progress (grade repeating, unique education positioning, and secondary school graduation), and social and emotional ability development and outcomes (Aos et al., 2004; Barnett, 2008; Belsky et al., 2007; Brownell et al., 2003; Network, 2005; Peisner-Feinberg et al., 2001; Sylva et al., 2004; Tietze et al., 1999).

There are multiple factors that can influence children’s participation in preschool education, including parents’ working situation and their opinion about preschool education. Families with higher income and education levels are more likely to let their children participate in a preschool education.

As awareness has increased, increasing numbers of children participate in early childhood education (Lombardi, 2008). In 1960, only 10% of children aged 3-4 years old participated in preschool education nationally. Based on the data from U.S. Department of Commerce, Census Bureau and American Community Survey (ACS), from 2010 through 2019, about 61% of 3-5 years old children enrolled in preschool education (*COE - Enrollment Rates of Young Children*, n.d.)<sup>1</sup>.

## 1.2 Preschool Social and Emotional Education with School Readiness

The ESSA promotes a well-rounded education that is superior to mathematics and English-based approach espoused by the previous ACT, No Child Left Behind (NCLB). The ESSA states that education outcomes should be assessed holistically in a way that covers both academic and social and emotional growth. It is important that schools, educators, families, and students themselves take accountability for success in school life. (Lehr & Osborn, 2005). Social and Emotional skills considered vital to children's success as they shift into the primary education period (Hemmeter et al., 2006).

Researchers and educator have defined key social emotional skills that children should prepare before entering, such as self-confidence, the ability to create positive social relationship (such as peers, teachers, and parents), concentration and perseverance when working on completing tasks, ability to effectively express emotions, ability to focus on the lecture in the classroom, and skills in dealing with interpersonal problems (Bowman et

---

<sup>1</sup> Data based on sample surveys of the entire population residing within the United States, including those living in group quarters (e.g., shelters, healthcare facilities, or correctional facilities)



al., n.d.; National Research Council, 2000). Self-regulation, emotional understanding, social trouble resolving, and social-emotional actions all contribute to good classroom management and succeed in academic climate (Denham et al., 2014).

Preschool education is the first stage of socialization training for children. Educators and researchers have emphasized the importance of social and emotional skills in preschool education. Social and emotional competency is an important indicator of school readiness (Berk, 2014). Children's social and emotional development outcomes in the preschool years could benefit children's overall development and ensure they are well prepared to enter school (S. M. Jones & Bouffard, 2012; McClelland et al., 2017). Social and emotional skills in preschool directly influence children's academic achievement in kindergarten (Torres et al., 2015), and positive preschool relationships are associated with higher adjustment and achievement in kindergarten (Bagdi & Vacca, 2005).

In existing studies on choice studies, researchers tend to position social and emotional development as being of the same important as cognitive ability for preschool education. However, over the years, social and emotional development education has been placed in a secondary position, behind cognitive ability education. According to the National Academy of Sciences, only 40% of kindergarten-age children had prepared social-emotional skills necessary to succeed, whereas 60% of children entered school with the cognitive abilities needed for success." (Yates et al., 2008).

Currently, all fifty states have released social and emotional development standards, including for preschool social and emotional education, and nine of them also have a K-12 social and emotional education requirement. Since the emergence of

conception of social and emotional learning, schools have paid more attention to social and emotional education. Social and emotional training programs have been established in thousands of schools, and researchers have conducted more than 500 evaluations of the various types of social and emotional learning programs. Many school districts are opening offices dedicated to social, emotional, and academic development. Social and emotional training coaches are working within schools, and social and emotional teaching is being incorporated into teacher preparation programs. Several states and national initiatives support students social, emotional, and academic growth.

### 1.3 Importance In Measurement of Social and Emotional Competency

Under the ESSA, non-cognitive indicators have been included in the accountability system. The need to measure non-cognitive indicators keeps increasing. Social and emotional development is a latent variable, which cannot be observed and scored in the same way one could evaluate writing or math ability.

To understand children's strengths and fulfill children's needs in emotional and social ability development, screening and measuring children's social and emotional competency are a critical step. Understanding each child's social emotional development from a practical standpoint aids practitioners in identifying a child's strengths and needs, as well as how these evolve over time. Malti et al. (2018) believe that knowing a child's social-emotional development is important because it can help prevent problems and lead to proactive interventions to support children's needs.

Correctly measuring children's emotional and social ability could help schools to adjust teaching content and parents to adjust the way of teaching their children, which

would help build a good relationship with children, prevent and intervene in potential mental problems, and correct existing behavior issues. There are multiple ways to accomplish the assessment, such as observation, behavior rating scales, interviewing, self-report procedure, projective expressive techniques, and sociometric methods (F. M. Gresham & Elliott, 1984). Out of these, observation and behavior rating scales are widely applied because they are easy to implement and low-cost.

There are multiple possible elements that contribute to the advancement of habit problems in children. At the individual child level, very early predictors of habit problems include temperamental problems, aggressiveness, language troubles, and also disobedience (Stormont, 2002). Household factors that are associated with problem behavior in children consist of maternal depression, severe parenting, stress domesticity events, limited social assistance and also family instability (Brooks-Gunn, 1997; Jones Harden et al., 2000; Spieker et al., 1999; Stormont, 1998).

Research shows that 10% to 15% preschool children typically have chronic mid- to moderate levels of behavioral problems; this percentage is even greater among children from families with financial difficulties (Huaqing Qi & Kaiser, 2003). National Early Interventional Study results showed that 10% to 40% infants and toddlers with a disability were identified as having behavior challenges (Katsiyannis et al., 2001). However, less than 10% children who engage in problem behavior receive appropriate support (Kazdin & Kendall, 1998).

Berk (2014) claimed that children's behaviors tend to stabilize at around 8 years of age, therefore the early childhood (5-8 years old) period is the appropriate time for schools and families to help children to build good ability in social and emotional management and

to correct problem behaviors. In terms of measuring social and emotional learning in this age range, children in this age group do not have a comprehensive ability to express their needs and opinions or to carry out self-response, and parents have difficulty to understand and explain the behavior of their children. A rating scale method can be used as a guide to help parents to understand their children, and a rating scale requires less time and can be more easily trained on teacher or parents to complete compared with other methods.

To better measure young children's social and emotional development, an effective and easily applicable instrument should be adopted to help schools, families, and communities assess the development status of students and provide timely and necessary support for children who need more help in the social and emotional learning process. The measurement result can be used directly to evaluate children's development in terms of social and emotional skills.

#### 1.4 Instrument Quality

A construct is a representation of something that does not exist as a visible dimension of habits, and the more abstract the construct, the harder it is to determine (Nunnally, 1978). When building an instrument, the construct definition range, criterion-related validity, and internal consistency should all be considered. These findings show construct validity, or the extent to which the scale measures what it claims to measure. According to Kerlinger (1986), construct validity serves as the link between theory and psychometric assessment, and it is crucial for the creation of high-quality measures (Schmitt & Klimoski, 1991).

Before applying an instrument, to verify it is functioning the way it intended and measuring what it should to is increasing important as the survey research method is widely adopted by researchers (K. D. Bradley et al., 2015). Validation is a process that involves gathering and analyzing data in order to determine an instrument's quality. To verify the latent variable's terminology, definition, attribute, and utility are essential. The inconsistency of these sides can influence the validity of an instrument (Wigelsworth et al., 2010). To assess the validity of quantitative instruments, a variety of statistical tests and measurements are used, with pilot testing being the most common.

Reliability can be regarded as the reproduction of an instrument. Test-retest and internal consistency method are usually applied to test the consistency of an instrument. For social and emotional ability measurement, a reliable result is essential to provide evidence about the strengths of diverse groups of children within a classroom (or the entire school) in ways that might not be readily discernable without a strong data-driven approach (Malti et al., 2016). Measurement results can also assist families in identifying the abilities and assist a kid needs to participate in daily activities in the home and community. (Hemmeter et al., 2006).

Education psychologists are increasingly likely to be involved in the measurement of social and emotional skills (Wigelsworth et al., 2010). What is meant by social and emotional ability and how they are best measured concerns by both researchers and educators. A systemic and effective validation of a social and emotional measurement scale is important for researchers to collect and use data to complete analyses, and to reference the results to assess students and program performance.

## 1.5 Education Performance Assessment

Education assessment is a systematic process of gathering and analyzing data from various sources to gain a thorough understanding of what students know, understand, and are able to do with their knowledge as a result of their educational experiences. The assessment results can be used to improve subsequent instruction and to enhance student learning experiences (Huba & Freed, 2000). There are multiple sources of evidence that can be used to assess preschool education programs. This includes observation results, records of implementation, checklists/tests, and development reports. The assessment process includes defining, selecting, designing, collecting, analyzing, interpreting, and using information to improve students' learning and development.

Education assessment is a critical component of education in multiple ways: assessment result is a good tool to tell the program result as it has a direction function to a program; it can guide the program from target set, education content choice, and instruction method design; assessment result can help educators identify the deficit or flaw in the program, to help them clear about factors that may influence education results, to make adjustments of their work, and to provide a high-quality support to their students; assessment results can also be used as a standard to make a judgement about teachers' work performance and students' pass-fail decision.

## 1.6 Purpose

Children's social and emotional development situation is an abstract construct. Measurement of children's social and emotional development requires higher level of accuracy. This inquiry calls for a well-researched and well-developed survey instrument to

help researchers to obtain accurate measure of constructs of children's social and emotional development. Applications of a well-developed measurement instrument can help to minimize measurement errors, improve data quality, and improve the usability of the data collected. Instrument quality assessment procedure provides evidence about the quality of the instrument. High-quality instruments can provide more robust results for educators and researchers.

There are a range of measurements that researchers and educators generally apply to children of different ages. These include the Social Skills Rating System (SSRS) (F. Gresham & Elliott, 1990), Preschool and Kindergarten Behavior Scale (PKB-2) (Merrell, 1994), Early Development Instrument (EDI) (Janus & Offord, 2007), etc. Current popular social and emotional learning instrument tends to use a checklist with a great number of questions to find out children with problem behaviors.

However, instruments with a great number of items will consume too much time to complete. For preschool children, who do not have self-report ability, all of items should be completed by their teachers, which would be a big burden for them. Additionally, Schmitt & Stuits (1985) suggest that shortening a measure can effectively minimize response biases caused by boredom or fatigue. While research in this area continues to advance, a simpler, less time-consuming, fewer items number, and easier-applied instruments are required by schools and families. Cost-efficacy needs to be considered when creating an instrument.

For family, social and emotional measurement results can assist parents to understand their children, and to provide support for helping children's development.

For program planners, there is a clear need for a standard tool for curricula or program planning. Social and emotional measurement scale can be utilized to collect data, and the measurement result can provide information to guide social and emotional learning program development. A well-regarded social and emotional learning interventions often include customized instruments (cover a comprehensive or specific components of social and emotional learning) for providing reliable assessment evidence which is part of the program evaluation.

In addition, existing scales are generally assessed under the methods of the Classic Test Theory (CTT). With CTT theory, the quality of an instrument is measured through the true score and expected score. However, CTT method can only provide overall assessment results through the total score, and so this method lacks investigation of each item, and sample can influence the result also (K. D. Bradley et al., 2015). The Rasch model (Rasch, 1960), which can be applied to deal with these assessing limitations. Especially instruments with a small number of items, making sure each item contributes to the measure construct is important. Instruments with a small number of items have multiple inherent advantages compared with instruments with a larger number of items. For example, they are less time consuming, easier to complete, and save money.

Few studies focus on assessment and improvement quality of social and emotional instruments for preschool children group. This study assesses an instrument performance to provide an advanced method to assess performance of a non-self-reported instrument which is designed to measure preschool children's social and emotional competency. Additionally, this study applied on equating design under Rasch theory in assessing the performance of a preschool social and emotional education program based on repeated



measure results from two time points. A reliable assessment result can be used to determine the extent to which the education program achieves its goals for children's social and emotional learning; and can guide revision at the program and also improve pedagogy to enhance children's learning experience.

The federal government is not involved in regulating preschool education programs. Each state has their own standards and regulations of preschool education. Assessing the quality of preschool education programs is important for family and program planners. Structural evaluation is generally more commonly used (such as teacher-child ratio) than assessment results. However, education assessment results can provide direct evidence about the quality of a preschool program. This situation happens because creating a specific program instrument is generally costly and time-consuming and norm-referenced (Evergreen & Coryn, 2012).

The purpose of this study is to assess a non-self-reported, small item-number item instrument based on Collaborative for Academic, Social, and Emotional Learning (CASEL)'s five components framework and North Carolina state's preschool education standard. The Rasch assessment result can guide future instrument design in this field. Additionally, the instrument validation result will be applied to modify the instrument and assess the social and emotional preschool intervention program within the measurement result from two time points. The stacking method will be applied to equate two times assessment result, and to provide researchers and educators with a interpretable data-driven approach in assessing of change (B. D. Wright, 2003).

## 1.7 Research Questions

The purpose of this study was to assess the Feeling Friend (TFF) instrument performance in measuring pre-school children's social and emotional development based on the CASEL's framework and North Carolina states' education standard. This instrument was designed to provide a checklist for schools and families to set a milestone of children's social and emotional learning process, and to help program planners and families to evaluate the performance of the social and emotional learning intervention program.

1. How well does the instrument measure the preschool children's social and emotional learning situation? This question is evaluated by testing the Rasch unidimensionality assumption and three reliability estimates (item and person reliability indices and separation indices). Item content validity will be applied to group items, to make sure all items fall into the official social and emotional framework, and make sure all items contribute to one assessment construct.

2. How well does each item contribute to the social and emotional construct measurement (how well does each item fit to the Rasch model)? This question will be evaluated by infit and outfit statistics of item fit and item difficulty of this instrument.

3. How well does the category design of the social and emotional instrument fit into the Rasch model? To evaluate the category design, the Rasch category threshold will be applied to test distance between categories of each item.

4. How well does the Stacking Comparison method assess children's social and emotional learning improvement after the training? To answer this question, the stacking method will be used to equate the pre- and post- assessment results to assess the result of the social and emotional learning program. Additionally, a T-test will be used to

test the statistical significance of the change between pre- and post- emotional assessment results.

## 1.8 Overview of Subsequent Dissertation Chapters

This chapter serves as an introduction, the first chapter out of five chapters of this dissertation. Chapter 2 reviews social and emotional education system background of United States, including social and emotional learning definition, social and emotional development history, and existing policies and standards related to the social and emotional development education of preschool children. Specifically, both federal and local (North Carolina)'s policies and standards will be discussed. How social and emotional learning outcomes affect children also will be discussed in this chapter. In addition, a brief introduction of currently widely-used social and emotional learning scales will be given. How social and emotional learning program can help to improve children's social and emotional learning experience will also be dealt with. Additionally, chapter 2 introduces the Rasch measurement theory and equating method that which can build link between two-time points assessment result. Chapter 3 details the research design procedures, instrument information, participant information, and applied Rasch model in assessing the instrument and assessing the intervention of social and emotional learning program performance. Chapter 4 presents the results of analyses, which includes descriptive statistics of participants, dimensionality, item fit, item difficulty, DIF, category thresholds, and stacking comparison result. These results of analyses will be utilized to assess the TFF instrument and the social and emotional learning program. Chapter 5 is the conclusions and limitations related to the instrument assessment, the implications of the research, and

this chapter also suggests recommendations for future research related to social and emotional education.

## CHAPTER 2. LITERATURE REVIEW

This chapter reviews the literature regarding the background of social and emotional education, the social emotional learning theory, the development of social and emotional education history, the social and emotional definition, the policy and standard of preschool readiness and social and emotional education, social and emotional competency effect in children's development, and discussion about currently widely-used-scales for social and emotional measurement.

Additionally, this chapter reviews literature related to the Rasch model, specifically the Rasch partial credit model, and the advantage in using the Rasch model to assess an instrument which is designed to measure unobserved constructs, instead of the Classical Test Theory method. Additionally, the reason for applying an equating method in comparison to repeatable measurement for assessment and its advantaging are also discussed in this chapter.

### 2.1 Social and Emotional Learning Background

#### 2.1.1 Social and Emotional Definition

Social and emotional competence is a broad concept with multiple facets; there are couples of positions under this term (Cohen, 2006). Social and Emotional competence can generally be separated into two parts: emotional competency and social cognition. Emotional competence is defined as the capacity to express oneself and others' emotions purposefully, fully understand one's own emotions and those of others, and regulate feelings and experiences when necessary (Denham, 2006). Social cognition or social

perceptions refer to how one perceives, understands, and empathizes with the emotions of others (Warren et al., 2008).

In the United States, the National Conference of State Legislatures (NCSL) gives the following definition of social and emotional learning:

*“Social and emotional learning refers to a wide range of skills, attitudes, and behaviors that can affect student success in school and life. Consider the skills not necessarily measured by tests: critical thinking, emotion management, conflict resolution, decision making, teamwork. While unable to traditionally quantify, these can round out student education and impact academic success, employability, self-esteem, relationships, and civic and community engagement”* (D. E. Jones et al., 2015).

Collaborative for Academic, Social, and Emotional Learning (CASEL) is a leading research-based organization. CASEL’s primary focus has been to synthesize empirical finding and theoretical developments to foster progress in Social and Emotional learning. Social and emotional learning is currently in the practical period, as an increasing number of schools, districts, and states have implement coordinated social and emotional education programming in their education system (Cohen, 2006).

By integrating social and emotional skills into general education, CASEL has built the social and emotional education frame. CASEL social and emotional education framework is comprised of self-awareness, self-management, social awareness,

relationship skills, and responsible decision-making (Schultz et al., 2011). The Interactive CASEL Wheel shows at figure 2.1 below:<sup>2</sup>

In December 2020, the Head CEO of CASEL updated the SEL definition:

*“Social and emotional learning (SEL) is an integral part of education and human development. SEL is the process through which all young people and adults acquire and apply the knowledge, skills, and attitudes to develop healthy identities, manage emotions and achieve personal and collective goals, feel and show empathy for others, establish and maintain supportive relationships, and make responsible and caring decisions.”*

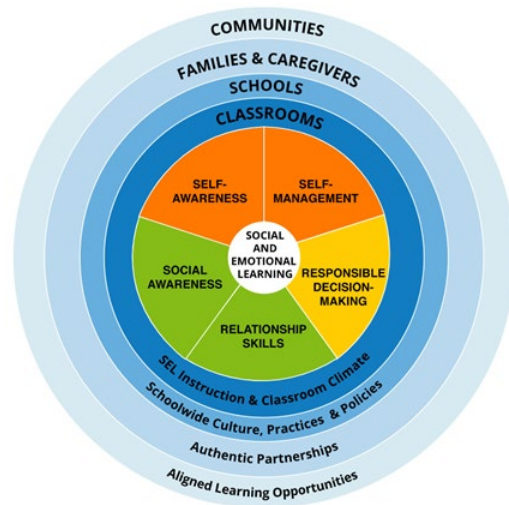


Figure 2.1: Interactive CASEL Wheel

---

<sup>2</sup> **Self- awareness:** the ability to understand one’s own emotions, thoughts, and values and how they influence behavior across contexts.

**Self- Regulation:** the ability to manage one’s emotions, thoughts, and behaviors effectively in different situation and to achieve goals and aspirations.

**Social – Awareness:** the ability to understand the perspective of and empathize with others including those from diverse backgrounds, cultures, and contexts.

**Relationship -skills:** the ability to establish and maintain healthy and supportive relationship and to effectively navigate setting with diverse individuals and group.

**Responsible Decision Making:** the ability to make caring and constructive choice about personal behavior and social interactions across diverse situations.

The Center on the Social Emotional Foundations for Early Learning (DSEFEL) defines social-emotional development as ability to build close relationship with adults and peers, from birth to age 5; to recognize, express, and manage emotions appropriately under a social environment; and to learn about the environment and family within the context of a diversity culture (Center on the Social Emotional Foundations for Early Learning, 2008).

### 2.1.2 Social and Emotional Education History

The social and emotional education theory concept can be traced to the philosopher Plato, who articulated his education philosophy in *The Republic*. He posited that the purpose of education is to develop children to be dutiful citizens. Regardless of the career path they choose, students need to know how to interact with others. In Plato's concept, social and emotional competency plays a critical role in living in society. Plato also affirmed the importance of early education. He believed early education a child receives influences the rest of his life.

In the United States, James Comer, the professor of child psychiatry at Yale University, developed the modern concept of social and emotional education (Edutopia, 2011). He believed that a child's psychosocial development could influence their academic achievement. In 1969, he and his colleagues created the Comer School Development Program (SDP), which is also known as the Comer Process or Comer Model. Its goal was



to improve the educational opportunities for ethnic minority youth. The SDP model shows at figure 2.2 below<sup>3</sup>.

The SDP model consisted of the process of planning, implementing, and assessing training program. It was designed to fulfill the needs of children's emotional development, such as confidence, respect, and positive interpersonal relationships at school and at home. Rewarding positive behavior will motivate children to continue positive behavior learning. Therefore, the SDP requires both schools and families to get involved in this training process. The SDP was implemented in multiple schools to train students how to correctly respond in a different situation, how to express themselves concisely, and how to interact with others (Anson et al., 1991). As an early-stage social and emotional learning program, the SDP points out the family and school corporate model. Researchers believe a child's social and emotional development is strongly influenced by teachers and parents (Jennings & Greenberg, 2009; Ozer, 2007).

---

<sup>3</sup> The SDP Model consists of three mechanisms (School Planning and Management Team, Mental health Team, and Parent/Family Team); three operations (Comprehensive School Plan, Staff Development Plan, and monitoring and assessment), and three guiding principles (collaboration, consensus decision making, and no-fault problem solving) (Lunenburg, 2011).

The Yale School Development Program Theory of Change

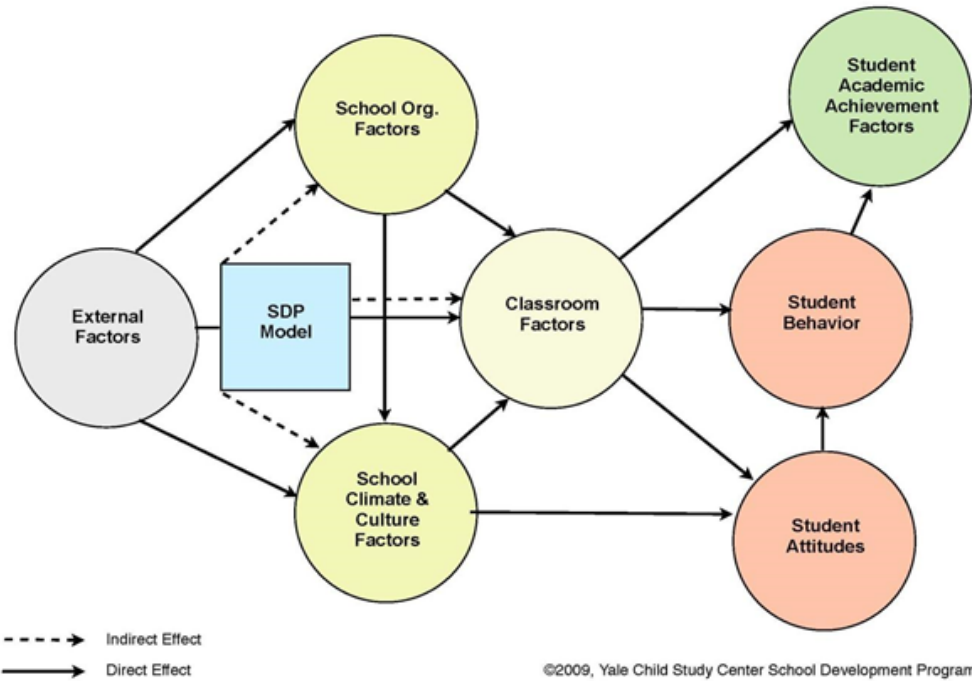


Figure 2.2: Comer SDP Model

The School-Based Promotion of Social Competence (1992) was established Along with support of the W.T Grant Consortium (Cummings & Haggerty, 1997). The purpose of this program is to set a standard to assess children’s development and encourage the family to get involved in preventing children’s potential behavior problems. In 1996, the program released a list of skills for parents to focus on when raising their children. The listed skills included compliance, following rules, social behavior skills (making friends), social and general problem-solving skills, coping with anger, self-directed academic behavior skills, understanding and expressing feelings, thinking helpful thoughts, and self-esteem (Hawkins et al., 1992 P.136).

A framework for integrating social and emotional learning into schools was developed by a group of well-known youth development and school-based preventive

experts. The group described these emotional competencies as "identifying and labeling feelings, expressing feelings, assessing the intensity of feelings, managing feelings, delaying gratification, controlling impulses, and reducing stress" (Hawkins et al., 1992 P.136).

In 1994, as the idea of social and emotional learning was gaining traction, the organization CASEL was founded under its original name, Collaborative to Advance Social and Emotional Learning. Among the scholars at this meeting were experts from multiple fields (e.g., emotional intelligence, child development, prevention science, bullying prevention, and public health) (Edutopia, 2011). Scholars were getting together to work on various projects related to children's social and emotional development, such as promoting healthy choices, school-community interactions, and generally responsible behavior in order to avoid violence and drug use in schools (Edutopia, 2011). The work *Promoting Social and Emotional Learning: Guidelines for Educators*, which was published by ASCD in 1997 and founded and defined the area, was coauthored by nine CASEL participants (Hargreaves, 1997).

### 2.1.3 Policy and Standard

The release of ESSA remedied the No Child Left Behind (NCLB)'s limitation of heavily focusing on math and writing scores. Unlike the NCLB, the ESSA also considered children's social and emotional learning as an indicator of student success. Under the ESSA, the goal of education is not only to improve students' test scores, but also to teach them the skills to live a happy and healthy life in society. The ESSA promotes a well-rounded education that is superior to the math and English-based approach espoused by the NCLB.

It states that education outcomes should be assessed holistically in a way that covers both academic and social and emotional ability growth. Schools, educators, families, and students must all take accountability for students' success (Lehr & Osborn, 2005). Moreover, Recent policy developments and the introduction of a new national strategy (the social and emotional Aspects of Learning program) have re-emphasized the importance of social and emotional skills in educational context (Wigelsworth et al., 2010).

As the ESSA has expanded the student achievement definition, "nonacademic" measures such as student involvement and school safety have been involved as part of student success. Social and emotional development is an indicator also. In response to the new act, the \$123 billion infusion to K-12 education in the American Rescue Plan (ARP) Act of 2021 represents an unprecedented opportunity to invest in social and emotional learning. The US Secretary of Education, Miguel Cardona, claims that "we need to address inequities in education, and we need better pathways to success . . . [which requires recognizing that students' social and emotional learning] is just as important as their academic growth]."

Under the ESSA, states have flexibility to set their own social and emotional learning accountability system, which includes definitions of social and emotional learning and determines what, when, where, and how social and emotional learning works in the classroom (Malti et al., 2018). Currently ten states have released legislation related to social and emotional learning, while all 50 states have released standards about preschool social and emotional learning. Six states have preschool and early elementary social and emotional learning standards, and nine states have preschool and K-12 social and emotional learning standards.

In North Carolina, the state that The Feeling Friend (TFF) instrument was primary designed for the goal of NC PK-12 in social and emotional learning: “Build and/or align state and local infrastructure to support social and emotional learning; Identify and promote statewide use of best practices for adult and student social emotional learning; Track, assess, and share advancements in social and emotional learning across the state.” NC released social and emotional development standards for Early Child education in 2013, which includes three components of social and emotional learning education: Sense of Self, Sense of Self with others, and Learning about Feelings (the North Carolina Foundations).

#### 2.1.4 social and emotional learning Accountability

**School:** School climate has been considered as the foundation for developing students’ social and emotional ability. Research shows school education is a strong tool to enhance children’s social and emotional competency. Schools with a supportive climate and well-designed instruction can improve students’ social and emotional learning experiences (Melnick et al., 2017). School also needs to apply measurement tools to assess the students’ social and emotional skill development.

**Teacher:** Teachers are the primary participator in students’ education. Teachers are responsible for implementing and following behavioral guidelines that promote intrinsic motivation, teaching children throughout conflict, encouraging student cooperation, and acting as a role model for polite and appropriate communication and prosocial behavior (Jennings & Greenberg, 2009). Teachers are also a direct and close observer of children’s daily activities.

**Family:** To enhance the effectiveness of social and emotional education, parents or major guardians need to keep consistent with the school education content.

**Students:** Students' engagement is another factor can influence the social and emotional learning outcomes.

#### 2.1.5 Importance of Social and Emotional Education

“Education should both provide immediate happiness and promote the long-term overall happiness of the community to a greater extent than any other kind of education” (Barrow, 2010 P.129). Education's objectives should be reframed to include not only classroom knowledge, but also social, emotional, and ethical capabilities (Cohen, 2006). The ultimate goal of education is to help children to develop skills to live happily in society

Previous studies showed that emotional and social ability plays an important role in children's prosocial orientation and reduction of problem behavior across development; Cohen (2006) suggests that social and emotional skills and behaviors are the foundation for a person to live in a democracy and achieve higher quality of life. Children with well-built social and emotional competency tend to have better academic performance, and happier academic lives, home lives, and daily lives in general. They tend to show lower aggression, develop more friendships, have better connections with parents and teachers, and engage in more interactions with peers. A person with higher social skills tends to show lower aggression, build positive peer relationship, and develop friendship earlier in life (Arnold et al., 2012; Landy, 2009; McCabe & Altamura, 2011; Rose-Krasnor, 1997).

To the opposite effect, lack of social and emotional competency could limit children's future academic achievement (Cohen, 2006). Children with lower social and emotional competence have difficulty in following the regulations and will probably not succeed in academics (McClelland 2006). Goleman (2006) more explicitly pointed out that those social skills are important factors for academic success.

Social and emotional education has been developed over the span of decades, and social and emotional learning research consists of multiple disciplines, such as preschool education, children health, and behavior research etc.

#### 2.1.6 Social And Emotional Scales

Recently, researchers kept working to improve the quality of behavior rating scales to provide more psychometric accuracy, which benefits educators and researchers who adopt these scales to measure social and emotional competency (Merrell, 2011). Moreover, teachers, parents, and other adults who participate in providing care to kids can use these instruments to understand children. They can also use instrument result to assess children's social and emotional ability change over time. Behavior rating scales can be used by teachers, parents, and other caregivers who are familiar with the kid and who use their knowledge of the child's behavior over time and in a variety of settings or situations. These behavior rating scales are generally designed to use a set of questions, statement, or items to cover a defined scope of typical children's behaviors, and based on the degree to which a child is aligned to the listed items, raters can give a score to an assigned child. Raters who use the behavior rating scale can have less time to complete the rating trained and use fewer time to complete (Wang et al., 2011).

The Social Skills Rating Scale (Gresham & Elliott, 1990) and Preschool and Kindergarten Behavior Scale (Merrell, 1994) currently are widely adopted by teachers to measure children at preschool and Kindergarten Level. These two scales are designed to find out whether a child exhibits problem behavior.

SSRS (Gresham & Elliott, 1990) was one of the most popular measures in assessing social and emotional development in preschool-age children (Merrell, 1994). Demaray et al. (1995) claim that the SSRS is a well-designed and easily applied instrument, which can provide the most comprehensive measurement result for assessing the social and emotional skills of preschool and school age children. SSRS includes three main components of social and emotional ability: Social Skills, Problem behaviors, and Academic competence. It is designed with a 3-point rating scale (never, often, very often). Researchers have rated the SSRS as a technically adequate, comprehensive, and acceptable instrument (Bracken et al., 1998; Demaray et al., 1995; Wittmer et al., 1996). However, since the SSRS is designed for children within a larger age range (both school age and preschool age), researchers concluded that the SSRS cannot provide adequate information when measuring preschool children's social and emotional ability (Fantuzzo et al., 1998).

Preschool and Kindergarten Behavior Scales (PKBS) was developed to fix the flaw of the SSRS, which was designed specifically for measuring social skills and social-emotional problem behaviors of preschool and kindergarten children from 3 to 6 years old. It includes 76 items, divided into two domains: Social Skills and Problem Behavior. Social Skills include three subscales: social cooperation, social interaction, and social independence. The Problem Behaviors Scale consists of two subscales: Externalizing problems and Internalizing problems. PKBS can be completed at both the school level and



home level. The PKBS is designed with 4-point scales (Never, rarely, sometimes, and often). Cordier et al. (2015) suggest that the PKBS is one of the scales with the most robust evidence of psychometric quality.

Devereux Student Strengths Assessment (DESSA) is another scale often adopted, and it collaborates between multiple domains to focus on measuring social-emotional competency. The DESSA scale includes 72 items and covers eight aspects of social-emotional competency, including self-awareness, social awareness, self-management, goal-directed behavior, relationship skills, personal responsibility, decision-making, and optimistic thinking. It is designed for children in kindergarten to 8<sup>th</sup> grade (Lebuffet al. 2009). The response providers can be parents, teachers, or people who provide care to children, but there is no self-report option (Nickerson, and Fishman 2009).

#### 2.1.7 Social and Emotional Development Intervention Program

Recently, social and emotional skills have been considered as important as cognitive ability in influencing children's academic performance. Social and emotional learning is a critical component in K-12 education. In addition, some scholars and researchers have advocated for more social and emotional learning in preschool education. Berk (2014) claimed that as young children's behavior tends to stabilize around the age of eight, teaching them effective communication and self-regulation before and during the early years of their education may reduce their risk for developing academic, behavioral, and mental health problems. Children with good ability in identifying and managing emotions in preschool will be well-prepared to enter school (S. M. Jones & Bouffard, 2012; McClelland et al., 2017; Schultz et al., 2011). Therefore, social and emotional health

classes are essential for young children to allow them to succeed in life, just like reading and writing classes. Lacking in social skills at the pre-school level can lead to poor educational adjustment, with long-term implications such as criminality, poor social functioning, substance addiction, violence against others, and dropping out of school. (Lyon et al., 1996; Rich et al., 2008).

Social and emotional skills training in the preschool period could help children in a variety of ways and have a long-term impact on their future. Early childhood is a vulnerable period for various areas of development, according to an intervention program undertaken in the preschool year (Bierman & Motamedi, 2015; Klibanoff et al., 2006; Raikes et al., 2006; Yoshikawa et al., 2013). Moreover, preschool intervention social and emotional learning programs have had long-lasting positive outcomes in enhancing the cognitive (Kautz et al., 2014; Reardon, 2018), social (Camilli et al., 2010), and economic (Kautz et al., 2014; Reardon, 2018) skills that impact children's future. Children themselves, family, and schools all benefit from social and emotional intervention programs (Hemmeter et al., 2006).

A direct intervention program could improve the social skills of both preschool and elementary-aged children (Caldarella et al., 2009; Kamps et al., 1995). As of now, social and emotional intervention programs are widely used to enhance children's development of social competence. Multiple researchers conducted meta-analyses that concluded that programs focused on fostering social and emotional skills in students from kindergarten to eighth grade could have beneficial outcomes, including skill competency, academic achievement, conduct problems, and emotional distress improvements. Murano et al. (2020) reviewed 48 studies, which included 15,498 preschool students, and the meta review

results showed that both targeted and universal social and emotional learning programs could improve children's social and emotional ability and reduce problem behavior. Payton et al. (2008) reviewed 324, 303 kindergarten through 8th grade students. Students involved in ongoing training of social and emotional skills showed a positive effect on the development of social and emotional skills, attitudes, and behaviors. Blewitt et al. (2019) concluded that universal social and emotional learning interventions in early childhood education and care setting could improve children's social and emotional development. The conclusion was based on 19 reviewed studies, which included 194 children.

The social and emotional Intervention programs can be classified into two categories based on the delivery method: the universal program and the targeted program. The universal social and emotional learning program will be delivered to all children. It usually provides a proactive and preventative approach that takes advantage of the preschool setting to enhance social-emotional capacity at the classroom level (Blewitt et al., 2019). The targeted intervention program is designed for selected children who are potentially experiencing social, emotional, or behavioral challenges. The program usually involves a parent training program to help children develop social and emotional abilities and decrease problematic behavior (Blewitt et al., 2019). Interventions such as these can result in improvements in parental and teacher interactions with children, and positively influence their learning process of social and emotional competency. It is critical that early intervention targets young children displaying disruptive behaviors, as the development of social skills may not be automatic for them (Webster-Stratton & Reid, 2003). Teachers play a pivotal role in the training phase of both social and emotional intervention methods. Several large-scale systematic reviews on preschool programs concluded that best practices

involved a strong emphasis on cultural relevance, teacher attitude and competence, and implementation (Bayer et al., 2009; McClelland et al., 2017; McLeod et al., 2017; White et al., 2017). Therefore, teacher training is an important competency of social and emotional training programs. Supporting early childhood teachers' concerns about behavior may have the unintended consequence of improving their ability to address other early learning outcomes to a greater extent, thus preparing young children for kindergarten success (Hemmeter et al., 2006).

## 2.2 Rasch

### 2.2.1 Rasch Measurement Theory

When we would like to measure an object, such as the weight of a bottle of water, the length of a sofa, or the volume of a cake, we could apply different instruments to measure them. The result of the measurement can be clearly read on the instrument. The unit is unified for all items and people understand the unit in the same way. For example, when we talk about the weather, we may say that today's temperature is 75°F, which is 5° F lower than yesterday. This information is very easily understandable by others because Fahrenheit as a measure of temperature is commonly used in the United States. However, to measure a person's ability, attitude, opinion, or feeling, such common instrument cannot be applied, because they are invisible and unobservable. For example, a university may want to know sense of belonging of students to the university, but belonging is about people's feelings and is therefore, unobservable. To measure these latent variables, the psychometric method could be applied to solve this problem.

The Rasch model is one of options can be applied when analyzing categorical data in sociological and educational studies which measure an unobserved latent variable (Rasch, 1960). Rasch measurement results can place a person's ability and item difficulty parameters estimation in the same ruler (B. Wright & Masters, 1982). The  $\theta$  is the person's ability, and the  $\delta$  is the item difficulty. The Rasch measure is based on the probability of the person correctly answering this item. In this formula, result will be transferred in logits (B. D. Wright & Linacre, 1989).

$$p(X_j = 1 | \theta, \delta_j) = \frac{e^{(\theta - \delta_j)}}{1 + e^{(\theta - \delta_j)}}$$

There are a couple of Estimation methods generally adopted by the Rasch analysis software, such as Joint Maximum Likelihood Estimate (JMLE), Conditional Maximum Likelihood Estimation (CMLE), and Modified Maximum Likelihood Estimation (MMLE). In this study, Winsteps was applied to analyze the data, which adopted JMLE. By maximum likelihood procedures, JMLE estimates item parameters using preliminary ability estimates and item responses for examinees. Estimating item parameters is then used for updating ability estimates by maximum likelihood procedures. This type of back-and-forth process is repeated until the item parameter estimates stabilize, using the updated ability estimates to update the estimates of the item parameters (J. M. Linacre & Wright, 2000).

Compared with other models which adjust variables to fit the data, such as Item Response Theory, Rasch model more concerned with developing the variable that is being used to measure the dimension of interest. Under this theory, the Rasch model is a good way to assess the instrument fit and to improve the precision of each item (*Item Response Theory | Columbia Public Health*, n.d.).

There are multiple advantages of assessing an instrument in the Rasch model rather than other CTT theory methods. The Rasch model provides the possibility to review and modify an instrument in a mathematical, quantitative way to provide evidence for measuring latent constructions such as opinion, feeling, and attitude. Rasch measures can also be used to correct bias brought by insufficient sample sizes or small item numbers. (J. M. Linacre, 1999; B. Wright, 1988). Moreover, the Rasch model is better able to structure an instrument, improve instrument quality, measure latent variables (social and emotional ability in this case), and estimate parameters of people and items at the same time.

The process of using the Rasch model to assess an instrument will be discussed in detail in the next chapter, Methodology, section 3.4.

### 2.2.2 Rasch model assumption

Rasch theory is a model prior, which requires a data to fit the model. Therefore, before applying the Rasch model, there are a couple of assumptions that need to be met (Gustafsson, 1980):

1. Unidimensionality
2. Local independence
3. Equal discrimination

The unidimensionality assumption is a priority among all three assumptions. It should be met before adopting the Rasch model to assess an instrument. The unidimensionality assumption is required for a specific instrument, as only one latent variable can be measured (J. Linacre, 2003).

Local independence includes three aspects: item local independence, which means that one item is independent of other items and will not influence the other items' performance; people independence, which means that one person's performance in this instrument will not be influenced by other people; and item and people independence, which means that in the Rasch model, estimates of person ability and item difficulty do not depend on each other. This means that the difficulty of the item does not vary across groups and the model's estimate for a group's response does not vary based on the difficulty of each item.

Equal discrimination. The Rasch model also implies that all items have equal discrimination between people with high and low ability. Additionally, Rasch model does not account for guessing.

### 2.2.3 Rasch Partial Credit Model

For analyzing polytomous option designs (with two or more ordinal options), both rating scale model (RSM) (Andrich, 1978) and Partial Credit model (PCM) can be applied. RSM can be applied to items that share the same rating scale, which is more about constraining the thresholds; PCM can be applied to analysis items in an instrument which does not share the same rating scale (Masters, 1982), which provide more flexible to investigation of the category structure design. In this study, the PCM model was applied. The equation of PCM is:

$$\ln\left(\frac{p_{n_i(x_i=k)}}{p_{n_i(x_i=k-1)}}\right) = \theta_n - \delta_i - \tau_{ik}$$

Where  $P_{ni(x_i=k)}$  is the probability that a person with ability  $\theta$  passing  $k$  number of thresholds on an item with difficulty  $\delta$  and a threshold set  $\tau_{ik}$  is the thresholds. The PCM can estimate person locations, item difficulties, and thresholds for each item. The result will be expressed on the logit scale.

### 2.3 Assess Results from Different Time Point

To understand the change overtime and to effectively compare the person measure from both pre- and post-form assessment results, an equating procedure needs to be applied to make the parameters estimated from both time assessments calibrated on a common scale.

According to Dorans (1990 P.3), “equating is a process of computing a statistical adjustment to the scores on one form of a test that will make them equivalent, in a particular way, to the scores on another form of the test”. After equating, a relationship between the different forms of tests will be established. The examinee’s ability estimate will be equal no matter what test form the person took. Equating is designed to adjust to differences in test form difficulty, with no difference in content (Kolen & Brennan, 2014; Yu & Osborn-Popp, 2019). Equating has more stringent requirements compared with linking, which compares a test score from one exam to another through equivalent transfer. Equating has been considered a more powerful form of linking (Holland, 2007).

To apply the equating procedure, the properties of equating should be followed:

1: Construct. This means the different forms of the test must evaluate the same construct. For example, if test 1 is designed to test verbal ability and test 2 is designed to test math ability, then these two tests could not be equated.



2: Equity. Same equity means there is no difference for examinees of the test given, and the two forms of test should have similar difficulty.

3: Symmetry (equating transformation). This requires the test results to be interchangeable.

4: Population invariance (equating function should be the same for all groups or subgroups).

5: Equal reliability (test forms should have the same reliability).

Meanwhile, currently researchers criticize these requirements, arguing that they are too tough to be apply (Dorans & Holland, 2000; Peabody, 2020). Equating may not be necessary for tests designed with equivalent construct and statistical specifications (Kolen & Brennan, 2014).

Equating methods could be categorized as Classical Test Theory (CTT) equating methods and Item Response Theory (IRT) equating methods based on the test theory. CTT methods and IRT equating methods were both developed for providing comparable test results on different forms of test in the same testing subject.

Rasch theory has multiple theoretical advantages compared with the CTT methods. The first one is the invariance attribute. That means the item parameters and person parameters are independent. In this theory, if the required equating data fit the assumption of the Rasch model, then it is possible to estimate the examinees' ability no matter which form of test the examinee took. The difficulty estimate of the item will not be influenced by the examinee's ability estimate, and the examinees' ability estimate will stay constant if the items' difficulty changed. No matter which form of test the examinee takes, the

examinee's ability estimate will be constant. The second advantage is that the examinee's ability and two test equated test forms can be placed on the same scale for comparison (Hambleton, 1989). Thirdly, since the item and person estimate are independent of each other, the Rasch methods can be applied to both parallel and non-parallel tests; there is no strict limitation for picking equating tests (Kolen, 1981). Lastly, because of the invariance attribute of the Rasch model, pre-equating methods are applicable. Pre-equating is a high-quality equating method with more flexibility for future comparison.

## 2.4 Summary

Social and emotional education has developed over a long period of time. Educators and researchers have demonstrated how social and emotional skills can influence young children in their primary school achievement and how they have a lasting influence in children's happiness. Schools, parents, and children are mainly involved in the social and emotional learning process.

To provide an understanding of preschool children and to provide timely support, researchers have developed a lot of instruments to measure their social and emotional ability. However, because the social and emotional competency's definition and range differ, previous instruments mix the age range in one scale, such as in the SSRS, which is designed for both preschool and school age children. Additionally, most of instruments designed with larger numbers of items to cover all possible aspects and set standards for children all over the country, which ignore the local differences within children.

Preschool education does not have a national standard, and each state has their own standard for preschool education, which fits the required readiness of state schools. To

develop a state-wide standard, which aligns with state requirements within a small number of items, is a good way to apply these standards through schools and families.

To assess a new instrument's performance and to assure the quality of an instrument, an accurate method is required by researchers. Compared with CTT assessment methods, the Rasch model has an advantage of reviewing each item making sure each item difficulty is appropriate and contributes to measuring the construct, other than the total score of the whole instrument. This provides researchers with more accurate evidence to modify an instrument.

## CHAPTER 3. METHODOLOGY

Chapter 3 details the instrument validation process in the Rasch partial credit model. Based on the validation result, the instrument will be reviewed and modified. An equating process (stacking) will be applied to assess the intervention program based on two times measurement results.

The Feeling Friend (TFF) instrument includes only 15 items, which is considered a small number of items compared with other instruments designed to assess similar constructs, such as PKBS, SSRS, and DESSA scales discussed in chapter 2. Based on the instrument design principle, each item should reveal at least one component of social and emotional competency to provide useful information for assessment. Assessing this instrument can also help with understanding how well the Rasch model works in validating children's social and emotional instrument (using a small number of items), and provide a comprehensive guide for using the Rasch model to guide researchers in improving an instrument.

The research process includes two parts: the first part is to assess the instrument using the data from pre-assessment results. Compared with the post-assessment, the pre-assessment includes more responses, which could provide more useful information. The Rasch model will be adopted to complete the assessment process. The second part is to make essential modifications to the instrument based on the Rasch validation results and to reorganize data. Then the equating process (staking) will be employed to complete the comparison of the pre- and post-assessment to assess the intervention program effort.

This chapter is organized by the research questions, and the research design is listed in Section 3.1. Sampling and instrument information will be provided in Section 3.2.

Following this, section 3.3 and 3.4 introduce the process to test the quality of this scale. Additionally, the equating comparison method (stacking) will be described to explain how the Rasch method works to assess two-time point assessment results.

### 3.1 Research questions

The purpose of this study was to assess the TFF instrument by measuring preschool children's social and emotional development based on the CASEL's framework and North Carolina states' preschool social and emotional education standard. This instrument is designed to provide a checklist for schools and families to set a milestone in children's social and emotional learning process, and to help program planners and families to assess the performance of the social and emotional learning intervention program. The study is guided by the following research questions:

1. How well does the instrument measure the preschool children's social and emotional learning situation?
2. How well does each item contribute to the social and emotional construct measurement (how well does each item fit to the Rasch model)?
3. How well does the category design of the social and emotional instrument fit into the Rasch model?
4. How well does the Stacking Comparison method assess children's social and emotional learning improvement after the training?

## 3.2 Sample

### 3.2.1 Participants

Data was collected from Portsmouth, Virginia, United States in 2019. There are twenty-three schools and institutions that form Portsmouth Public Schools (PPS): three high schools, three middle schools, thirteen elementary schools, three preschool centers and an alternative education center. PPS has over 2,100 staff members, and the division has about 13,000 students enrolled throughout the city. All students must be actively involved in their learning if Portsmouth Public Schools is to promote academic excellence and responsible citizenship.

In 2019, the Portsmouth school district provided training opportunities to teachers on implementing social and emotional learning curriculum content in their classrooms. Twenty-three teachers from three preschool education center sites: Churchland, Mt. Hermon, and Olive Branch participated in this training. After they got trained, they provided the social and emotional training content to their students in their classrooms. Children received thirteen weeks of social and emotional training in the classroom from these teachers.

The instrument of The Feeling Friends in School (TFF) was administered in both pre- and post-assessment. In total 336 children were assessed in the pre-assessment, and 289 children were assessed in the post-assessment.

All children were at the preschool level: 44.3% (149) of children in the pre-assessment were girls and 55.6% (187) of children were boys; 75.9% (255) of children were 4 years old, and 24.1% (81) of children were 5 years old. 45.3% (131) of children in the post-assessment were girls and 54.7% (149) of children in the post-assessment were

boys. In the post-assessment, 25% (72) of children were 4 years old and 75% (217) of children were 5 years old. The assessments were administered before implementing a social and emotional training program called The Feeling Friends and after completing the program at the end of the school year. The TFF is a universal 13-week training program delivered to all children (not only delivered to targeted children). The demographic distribution is shown in the table below.

All responses were answered by the children's teacher since children are too young to respond to the items. Teachers received eight hours of training on how to implement the curriculum content in the classroom prior to completing the assessment. Table 3.1 below shows the demographic information of child participants from both pre- and post-assessment.

Table 3.1: Sample Distribution

**Pre-Assessment**

<b>Demographic Characteristic</b>	<b>Full Valid Sample</b>		<b>Demographic Characteristic</b>	<b>Full Valid Sample</b>	
	<b>n</b>	<b>%</b>		<b>n</b>	<b>%</b>
<b>Respondent Gender</b>			<b>Race</b>		
Male	187	55.65%	African American	255	75.89%
Female	149	44.35%	Caucasian	43	12.80%
<b>Site</b>			Hispanic	6	1.79%
Churchland	91	27.16%	Asian	2	0.60%
Mt. Hermon	96	28.66%	Biracial	28	8.33%
Olive Branch	148	44.18%	Other	2	0.60%
<b>Primary Language</b>			<b>Age</b>		
English	332	99.10%	4 years old	255	75.89%
Non-English	3	0.90%	5 years old	81	24.11%

**Post- Assessment**

<b>Demographic Characteristic</b>	<b>Full Valid Sample</b>		<b>Demographic Characteristic</b>	<b>Full Valid Sample</b>	
	<b>n</b>	<b>%</b>		<b>n</b>	<b>%</b>
<b>Respondent Gender</b>			<b>Race</b>		
Male	158	54.67%	African American	219	75.78%
Female	131	45.33%	Caucasian	36	12.46%
<b>Site</b>			Hispanic	8	2.77%
Churchland	85	29.41%	Asian	1	0.35%
Mt. Hermon	81	28.03%	Biracial	20	6.92%
Olive Branch	123	42.56%	Other	5	1.73%
<b>Primary Language</b>			<b>Age</b>		
English	285	98.62%	4 years old	72	24.91%
Non-English	4	1.38%	5 years old	217	75.09%

3.3 The Feeling Friend Instrument

The Feeling Friend (TFF) assessment was developed based on National Head Start Standards and the North Carolina Foundation for Early Learning and Development Standards, which are designed to provide milestones for preschool children’s social and



emotional development. The assessment includes fifteen items about some daily actions or reactions under certain curriculums for children, and demographic questions about age, race, gender, primary language, and grade level. Teachers observe the frequency of how often children perform these listed actions to measure children's social and emotional competency, assessing how children recognize their own and others' emotions, how children deal with conflict with other peers, and how children interact with peers and adults. The instrument uses a rating scale from 1 to 5, which indicates Never, Rarely, Sometimes, Often, and Always. The raw score range of this instrument is from 15 to 75. No negative word item is included in the whole instrument. Thus, a higher score indicates children tend to perform with higher social and emotional competency, and a lower score indicates that children tend to have lower social and emotional competency. This measurement is targeted to assess children's social and emotional development but is not aimed to find students with a deficit in social behavior. The instrument measurement results can be used for guiding instruction and for assessing the learning result of the social-emotional training program also. The target population of this instrument is young children in the preschool and Kindergarten stages.

### 3.4 Instrument Construct Validity

#### 3.4.1 Content Validity

Before applying the Rasch validation process, the item content validity had been applied first. The CASEL framework and North Carolina Preschool social and emotional teaching standard content were applied to test the content of each item.

According to Nardi (2018, P. 62) “Content validity is a subjective way to understand how well a set of items measures the complexity of a concept or variable we are studying.” Content validity could provide evidence about instrument dimensionality. Making sure instrument content contributes to a single topic is important before applying the Rasch model validation process. The CASEL framework and North Carolina Pre-Kindergarten Emotional and Social Development (ESD) framework will be applied to categorize and sort items based on their similarity to construct definitions.

### 3.4.2 Rasch Assumption Test

As mentioned in the previous chapter, to apply the Rasch model validation and measurement, the prerequisite assumptions should be met. The unidimensionality required that all items in one instrument measure only one construct. The main variance among samples can be explained by one component, and residual variance cannot be concluded to another components.

The unidimensionality of this instrument will be assessed with a principal component analysis of residuals (PCAR). The PCAR methods examined the extent to which observation variance can be explained by the model. The dimensionality judgement is based on both items’ difficulty and people’s ability. Less contrast is preferred as it shows that the more variance is explained (J. M. Linacre, 2000), which also means low possibility that a secondary component exists. The eigenvalue in the first contrast being less than 2 is considered random noise. Based on Linacre's (2012) suggestion, the Rasch model explained more than 50% of the variance; the first contrast eigenvalue was less than 2 (the

equivalent of 2 items). An eigenvalue of less than 2 can be considered as the noise, which does not influence the dimensionality of this instrument; in the other word, less than two items are not enough to suggest another dimension exists. Linacre (2000) suggests that an eigenvalue greater than 3 indicates that the instrument is multidimensional; in this study, cut point 3 will be adopted.

The Winsteps PCAR output also provides the residual loadings for the first contrast. One cluster of items shows positive loading, and another cluster of items displays negative loading in the residuals. When loading items in positive and negative with a contrasting pattern, this pattern can remind the researcher of that their probability has multiple dimensions (which are) also included in this instrument.

If the PCAR test result does not meet the criteria above, then that may indicate to the researcher that there is likely more than one dimension and further review needs to be applied. However, that does not mean the instrument does not meet the Rasch model at all. The multifaceted construct also can create this result; therefore, researchers can make a decision about whether adjusting the instrument or not being based on the latent variable theoretical framework. In this study, based on the CASEL framework and the NC learning standard of social and emotional education, as there are more than two components included in the social and emotional competency framework, then the PCAR test may suggest that there is another dimension exist beside the primary factor.

### 3.4.3 Rasch Model in Transfer Ordinal Data to Equal Interval Data

In this instrument response design, the response categories range from 1 to 5, which indicates the different frequency of item-listed actions. Therefore, the raw score collected

from pre- and post-assessment is ordinal data, which is not equal interval, which means the researcher cannot tell the distance between 1 and 2, and the distance between options is not identical. Rasch analysis can help in developing ordered continuums in which latent variables are measured by an instrument that fulfilled the unidimensionality assumption of the Rasch model (Michell, 2014). The Rasch model uses items and persons to define a common interval scale conjointly (Coombs et al., 1970), whereas the logit scale is defined based on the probability of a particular person within an endorsed ability being able to select a rating category on items. Parameters of person and item estimation will be discussed detailed later.

#### 3.4.4 Rasch Partial Credit Model

In this instrument category design, each item was originally designed with five options: Never, Rarely, Sometimes, Often, and Always. The instrument adopted odd number categories design, which includes a middle point option. K. D. Bradley et al. (2011) claimed that the middle point design can lower the quality of the measurement. In this study, I would like to challenge the category design by using Rasch model to test if the category design is appropriate to meet the measurement purpose with Rasch-Andrich thresholds.

PCM allows each item to have a different category design (different number of options or different distance category threshold). Therefore, it provides researcher to test that if all items category designs are functioning as expected.

Another reason for adopting PCM is the sample size. In this study, in the pre-assessment, 366 children had been assessed; for some categories, there do not have enough

observed (less than 10) for some items; for testing each item category functioning, the PCM was applied.

#### 3.4.5 Instrument validity and reliability

The pre-assessment data will be used to assess the measurement. The Rasch validation process will be completed by Winsteps 5.1.7 (Linacre, 2021), a Rasch specialist application. Data are organized based on Winsteps's adapted format, and a Winsteps's control file has been created.

The Rasch model will provide multiple diagnoses for this instrument. Reliability, and separation indices will be applied to make sure the overall instrument can provide reliable and reproducible results in measuring preschool age children's social and emotional ability. The reliability index shows how reproducible the item difficulty order is for items for sample people. It is expected to be larger than .8 (J. M. Linacre, 2012b). A reliability score of less than .5 indicates that the difference among individuals is due to measurement error (J. M. Linacre, 2012b). The separation index is an indicator of how well the instrument can sort people into different levels of ability. It is expected to be equal to or greater than 2 (Kook & Varni, 2008). If the separation indices are smaller than 2, which indicates the instrument cannot successfully separate children into 2 different groups, it may indicate that the instrument cannot measure children's social and emotional ability (Kook & Varni, 2008).

#### 3.4.6 Individual Item Quality test

#### 3.4.6.1 Point- Measure Correlation

J. M. Linacre (2012b) states that the point-measure correlation can provide information about the relationship between one person's ability to respond to an item. Noticeably positive correlations are expected. A high correlation indicates the person with higher ability should have a higher rating on items. Reversely, a person who gets a higher rating on items should have a higher ability in the latent variable that the research intends to measure.

Negative correlations usually suggest that the item's responses contradict the latent variable's direction. If there had been negative connections, an item with a reversed meaning may have been included in this survey, and therefore researchers need to review the item or consider removing this item from the scale.

#### 3.4.6.2 Item Fit

The real data usually cannot fit the model perfectly, which is also true for the Rasch model. Understanding which item does not fit could provide evidence to modify and improve the quality of the instrument. Fit statistics are formed through summarized residuals of the observed responses from real data and the estimate expectations which assume data all fit the model to find misfit items or people (B. Wright & Masters, 1982). In the Rasch model, to test the fit of the scale, four statistics are usually reviewed: infit mean square (Infit-MNSQ), standardized Infit (Infit-ZSTD), outfit mean square (Outfit-MNSQ) and standardized Onfit (Outfit-ZSTD). Compare with Infit-MNSQ, Outfit-MNSQ is unweighted, thus Outfit-MNSQ statistic is sensitive to outliers due to guessing or

careless errors, and Infit-MNSQ is weighted, which based more on responses near a given difficulty (or person ability) (Boone et al., 2014b). Outfit is easier to identify and correct; for this reason, item outfit needs to be paid more attention to, instead of range. (Boone et al., 2014a). The Mean-square fit statistic is the amount of randomness in response to an item; it shows the distortion amount in the measurement. The generally accepted value of infit mean square and outfit mean square is between 0.6 to 1.4 (B. Wright, 1994) ; an item fit of less than 0.6 indicates an item is overly predictable and too easy to complete; and item fit larger than 1.4 indicates unproductive for construction of measurement; an item fit of greater than 2 which may create a degrading problem – the people actually with higher ability have a higher possibility of getting an incorrect score compared to people actually with lower ability (B. Wright, 1994). Infit-ZSTD and Outfit-ZSTD statistics use a t-test to answer the question about how well the data fit the model (Mueller & Bradley, 2009). Generally, the infit-ZSTD and outfit-ZSTD are expected to fall into the range between -2 to 2, if ZSTD value over this range indicates that the Mean square value is not likely to be due to chance, there may a flaw in the data fit (B. Wright, 1994).

Fit mean-square and standardized fit work together to provide evidence about how well each item works in collecting the information for measuring construct. Each item included in the instrument should contribute to measuring the latent variable. If an item has a misfit issue, researchers need to review it and make a decision of modifying or removing it from the instrument.

#### 3.4.7 Item and Person parameter estimation

The Rasch model can be used to estimate the parameters of person and item. The person parameter indicates the ability, and the item parameter indicates the difficulty of the item. The Rasch model estimates a person's ability and item difficulty on a common scale. The Rasch model estimates the parameter result, which will provide evidence about how well the instrument items test children with different levels of social and emotional competency. Additionally, two-time points parameters estimation can provide evidence about the children's improvement in their social and emotional skills.

As discussed before, the JMLE estimate was adopted for this study. Another advantage of Rasch parameters estimation with JMLE is the missing data, which does not influence the estimation of person and item. Person's ability parameter and item's difficulty parameters estimates will provide the answer about personal ability and item difficulty match and distribution.

#### 3.4.8 Differential Item Functioning (DIF)

Item bias of an instrument is another problem that may influence the instrument measurement quality. To make sure items measure people within different backgrounds in the same way is important for keeping the instrument fair. The differential item functioning (DIF) can help researchers to diagnose potential item bias of an instrument (Lord, 2012). Additionally, DIF can help in finding differences between subgroups in a particular item (Choi et al., 2006). The Winsteps investigates item bias by estimating an item's difficulty of subgroups and keeping other items' difficulties and person measures constant. In this instrument, children from different subgroups present different response probabilities in an item, which may indicate the item with bias in a particular subgroup. The Winsteps analysis



application provides multiple indices to recognize DIF in an instrument. In this study, Rasch DIF contrast, Rasch-Welch *t*-test, and Mantel DIF chi-square indices will be used to test the DIF.

The *t*-test is applied to evaluate the difference between two groups, and to answer whether there are statistical differences between any two groups. If the *t*-test result shows the statistical significance between subgroups, this may indicate the DIF exists. DIF difference present can be either interpreted by item design bias or a real difference in groups (Conrad et al., 2007). In the Winsteps DIF analysis with *t*-test, outliers are not included in the analysis, because items cannot separate people based on their endorsed ability (J. M. Linacre, 2012a).

Welch's *t*-test statistic has been adopted to test if there is any item bias in subgroups. Linacre (2012b) suggests that when the DIF score is more than .5 logit, *p* value less than .05, then the item needs further investigation.

#### 3.4.9 Category threshold

Rasch - Andrich threshold, also called the “step calibration” or “step difficulty” is usually represented with the Greek letter  $\tau$  (Tau). It is the point on the latent variable (relative to the item difficulty) where the possibility of being observed in category *j* equals the possibility of being observed in category *j*-1.  $F_j$  is the Rasch – Andrich threshold. In the PCM,  $F_j$  is defined differently over items (J. M. Linacre, 2012b).

$$\log_e (P_{nj} / P_{nj-1}) = B_n - D_i - F_j$$

As with Rasch's original dichotomous model,  $B_n$  is the person's ability, attitude, capability, etc.  $D_i$  is the item difficulty, challenge, impediment. The social and emotional items on the scale of the Feeling Friends have five thresholds, as the accompanying rating scale for each item is comprised of five options (1 - Never, 2 - Rarely, 3- Sometimes, 4 - Often, and 5 – Always). Each child can have only one of five response options.

The category order should be in accordance with the latent variable. If disordering is observed, that indicates some of the category's responses are too rarely observed. This implies that these options correspond to an interval on the latent variable that is not big enough. Andrich thresholds between response categories should fall into the range of 1.4 logits to 5 logits (J. M. Linacre, 1999).

### 3.5 Assessment of Children's social and emotional Competency Change Over Time

To effectively compare the person measure from both pre-and post-form instruments, an equating procedure needs to be applied to make the parameters estimated from both times assessments calibrate in a common scale. Equating is a method for adjusting the scores on one form of a test so that they are equivalent to the scores on a different form of that test, in a particular way (Dorans, 1990). After equating, a relationship between the different forms of test will be established. The examinee's ability estimate will be equal no matter which test form the person took. Equating is designed to adjust for differences in test form difficulty, with no difference in content (Kolen & Brennan, 2014; Yu & Osborn-Popp, 2019). Equating is the most powerful method with the most stringent requirement compared with other methods (Holland, 2007). To evaluate the program performance, a comparison of pre-and post-assessment results is needed. To put two times

assessment results on the same scale, equating is essential for measuring change in children's social and emotional learning outcome over time.

The instrument had been applied to both pre-and post-assessments, with the same construct, to measure the development of children's social and emotional competency, which ensures both measurements have the same difficulty and design. Furthermore, the children who were involved in the pre-assessment also participated in the post-assessment; therefore, the sample could be considered equivalent. Equating could be applied because the pre-and post-assessment shares the same measure construct, equivalence population, and measurement design (C. H. Yu & Popp, 2005).

There are a couple of decisions that need to make before applying the equating. First is the sample design method. Based on the raw data, a common-item design has been chosen. Common items are design required: some common items exist in two forms of test being given to two groups of examinees. The two groups of examinees' abilities do not need to be equivalent. Researchers claimed that the more common items, the fewer random equating mistakes (Budescu, 1985; Wingersky et al., 1987), and too few common items will lower the equating quality (Petersen et al., 1983). To apply the common item design, the common-item anchor should be at 20% of the total items (Cook & Eignor, 1991).

The second decision is parameter estimation. Concurrent (simultaneous) estimate parameters, which means that two tests' data will be put into one dataset, and items from two test forms and examinees' parameters, who participated in two different test forms, will be estimated together. The concurrent estimation method allows the calibration of all items from both test forms. With the concurrent calibration step, the response from both test forms will be linked automatically and calibrated on one scale. Hanson & Béguin

(2002)'s simulation study resulted from comparing the mean/sigma, mean/mean, Haebara, and used concurrent calibration methods to conclude that when data fit the IRT model, compared with other methods, the concurrent calibration procedures generated more accurate evidence than test characteristic curve approach. Additionally, mean/mean and mean/sigma methods were less accuracy than other methods (J. M. Linacre, 2012a).

Under the Rasch theory, the concurrent common-item equating design is also named the stacking comparison method (B. D. Wright, 2003). The stacking method was targeted to compare a person's ability to change over time, which fulfills our research requirement and tests the competency change pre- and post-participation in the program. To apply the stacking method, the data needs to be appended to the person measure for pre- and post- responses based on the picture shown below. Data is stacked vertically, so each person appears twice (pre, post) for each item as figure 3.1 shows below (B. D. Wright, 1996). This method places pre-and post-assessment performance estimates for each person along the same continuum and allows for the examination of changes in social and emotional ability on each scale from pre- to post-assessment (Cunningham & Bradley, 2010).

Since the TFF instrument was administrated twice without change, we can consider the total 15 items as common items. The concurrent estimate method will be applied by putting the two times assessment results into one dataset. In this situation, the item parameter and person's ability parameter from both tests will be estimated simultaneously; estimates from both assessments will be on the same scale automatically.

In the Stacking comparison, we estimate the pre-training and post-training children and items parameters simultaneously(Wingersky & Lord, 1984). With this method, we fix

the parameters of item difficulty and calibrate children's ability from both pre-and post-assessment. After equating, we will have one set of parameters of item difficulty and both pre-and post-assessment parameter estimates of children's ability on the same scale.

The stacking method may be questioned for potentially violating the local independence of the participants. However, we hypothesize that children's social and emotional ability has changed during eight weeks of training. Based on this hypothesis, we consider children who participated in the pre-and post-assessment as not the same group of sample, which still follows the independence rule of the Rasch model.

After equating two-time points assessment results, we will have item difficulty, children's ability before the training, and children's ability after the training estimated in a common ruler. By comparing the overall change in children's ability in pre-and post-assessment, we can evaluate the program's influence on children's social and emotional development. To test the change for the pre-and post-assessment, a *t*-test can be applied to compare the means of the pre- and post-tests parameters estimates.

The *t*-test is a static method that is usually applied to compare the difference in means between two groups. In this study, we will adopt the *t*-test to compare the pre-and post- children's ability differences to test after the training, if the children have statistically significantly improved in social and emotional competency. JMP software (JMP, 2018) will be used to complete the *t*-test in this study. We hypothesize that children's social and emotional competencies do not have significant differences before and after the training and set  $\alpha = .05$ . If the *p*-value is less than .05, then the result provides strong evidence to reject the null hypothesis, which suggests the training program can significantly help to improve children's social and emotional ability.

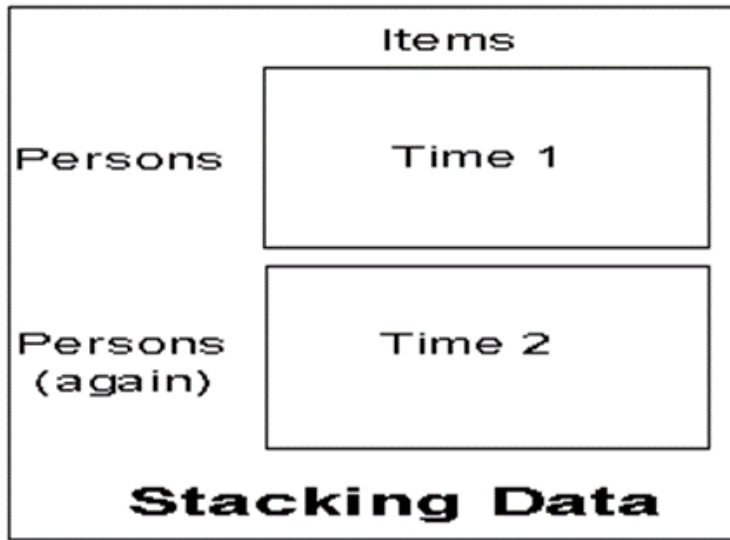


Figure 3.1: Stacking Data Structure

### 3.6 Wright Map (Variable Map) and KeyForm

#### 3.6.1 Wright Map (Variable Map)

The Wright map is an important and unique output of the Rasch model analysis. It provides results of visual parameter estimates: people parameter and item parameter estimates are placed on the same scale simultaneously. Each person and each item will be calibrated in this ruler, and items and people are located on the two sides of the vertical ruler. Generally, the left side of the map calibrates the person's ability to measure and the right side calibrates the distribution of item difficulty measure. Both people and items are calibrated in a qualitative order from low to high. The item located on the bottom of the ruler indicates that the item is easier than an item above it. The person located on the bottom of the ruler indicates the person has a lower estimate of the latent variable than a person on the above location. By comparing the location, researchers can tell how a particular item's difficulty meets an assigned person's ability. It also helps to compare an item with a person,

if one item locates above a person, which indicates this item is hard for this person to get a higher rating. The variable maps also mark the means of item difficulty and a person's ability. "M" on the map indicates mean; "S" indicates standard deviation; "T" means two standard deviations. The number marked on the ruler in logits aligns with the parameter estimate.

The Wright map can provide answers about how well items' difficulties match people's abilities. K. Bradley & Sampson (2005) suggest:

"In a well-targeted assessment, mean item and person measures should be approximately equivalent, the difficulty measures of the item should span at least the width of the ability measures of the persons taking the assessment and the items should be distributed such that they accurately measure all persons taking the test. When a group of persons falls in a space between item placements on the ruler, it is comparable to measuring the length of the object; one could report the length of the object is between certain units, but could not report the precise length" (P.13).

The Wright map can also provide the change in people's abilities over time. Under the stacking comparison curriculum, the research could give the answer about how a person's ability changes over time based on the person's two time-calibrated locations on the ruler, which could provide a non-parameter result for comparison. A distance between two times locations in the hierarchy illustrates how the person's ability changes over time. If the post-assessment calibrates above the pre-assessment, that indicates the person has an improvement over time, and a larger distance means a bigger improvement. On the contrary, the person has a decrease in the ability over time.

### 3.6.2 Keyform

Keyform is another output of Winsteps. Compare with Wright map which shows the item and person measures. For items, this means the average of the item thresholds. However, the Keyform can present each threshold map across all of the items. The Keyform can provide more information for polytomous data analysis. The x-axis represents people's ability, and the y-axis presents items' difficulty. Additionally, each item expected rating categories pattern to be also displayed. Categories pattern usually step-like other than column-like, because items normally have different difficulty levels (Peabody et al., 2017).

### 3.7 Chapter Summary

Chapter 3 detailed described how the Rasch partial credit model is applied in two ways. First, the Rasch model is applied to assess the measurement performance of the instrument and improve the instrument. The second is to use the equating method (stacking) to compare the two times assessment results to assess the training program performance.

Chapter 3 mainly presented the procedure involved in validating the instrument. PCAR and content validity methods were applied in both qualitative and quantitative ways to ensure the unidimensionality of the instrument. Then, the Winsteps software was adopted to complete the Rasch analysis. To gain the whole picture of the instrument quality, we combine instrument reliability and validity analysis, reviewing the item quality of each item, item and person parameter analysis, DIF analysis of potential items bias, and category threshold analysis results. The result had been referenced to modify and improve the quality of this instrument. The essential modification was applied to instrument design and



original data before the stacking comparison procedure. Then the Rasch analysis was used to identify if the items bias included in this instrument.

The equating method (stacking) has been described in detail including how to use the result to compare the change over time, the advantage of this equating method, and why this method can be applied in this study.

Additionally, methods of examining the Rasch measurement result have been discussed. This includes Rasch assumption examination, reliability and separation indices, person and item parameter estimation, fit statistics, Category thresholds examination, and DIF testing, which have also been explained.

## CHAPTER 4. RESULTS

The purpose of this study is to assess a preschool children's social and emotional instrument by using the Rasch model diagnosis and to use diagnosis results to improve the measurement quality of the instrument. The equating method (stacking) was adopted to compare the pre-and post-assessment results to assess the training program effect. Due to the limitation of the Classic Test Theory (CTT) theory which has been discussed in Chapter 2, an advanced method is needed to further investigate this instrument, which including verifying dimensionality, evaluating overall instrument quality (reliability and validity), investigating single item quality, item bias, and review the category design. In this chapter, data analysis results of The Feeling Friend instrument validation will be presented based on the research questions. The diagnosis processes and results include: unidimensionality, separation, reliability, item fit, DIF, and Category threshold. Subsequently, the data was modified based on the diagnosis results, then the stacking comparison equating method was applied to assess the training program performance in changing preschool children's social and emotional skills.

### 4.1 Content Validity

The TFF assessment items include how children recognize their own and others' emotions, how children deal with conflict with other peers, and how children interact with others (adults, friends, and parents).

Collaborative for academic, social, and emotional learning (CASEL) and pre-Kindergarten Emotional and Social Development (ESD) standards was used to identify

five interrelated sets of cognitive, affective, and behavior competency of group items, and to make sure items' contents followed the requirements of CASEL and ESD.

The TFF assessment includes 15 daily activities and behaviors to help teachers observe children's actions. Based on the content of items' test, they have been grouped based on CASEL and NC's ESD standards. The content validation result showed all 15 items' contents are fallen in the range of CASEL's five components and NC's ESD standards. Some items cover more than one component, and all items are grouped based on their majority measure purpose. The group result shows in table 4.1 below.

Table 4.1: Item Group based on CASEL and NC social and emotional preschool learning standard.

#	Item	CASEL	NC_ESD
<b>Item 2</b>	Can talk about own feelings	Self-awareness	developing a sense of self
<b>Item 7</b>	Demonstrates an understanding of their own uniqueness. (Ex: talents, interests, preferences, or culture)	Self-awareness	developing a sense of self
<b>Item 9</b>	Uses various coping techniques to manage emotions with the support of an adult.	Self-management	the learning about the feeling
<b>Item 10</b>	Transitions between activities without getting upset.	Self-management	the learning about the feeling
<b>Item 1</b>	Makes empathetic statements to other children.	Social Awareness	the learning about feelings
<b>Item 3</b>	Identifies basic emotions in books or photographs.	Social Awareness	the learning about feelings
<b>Item 5</b>	Communicates feelings in manners that are appropriate to the situation.	Social Awareness	the learning about feelings
<b>Item 4</b>	Can describe the impact their behavior has on others.	Social Awareness	developing a sense of self with others
<b>Item 6</b>	Engages in some positive interactions with strangers. (Ex: parent volunteers)	Relationship skills	developing a sense of self with others
<b>Item 11</b>	Able to follow classroom rules.	Relationship skills	developing a sense of self with others
<b>Item 12</b>	Seeks help from adults when problems arise. (Something to do together, joining an existing activity, or sharing a toy)	Relationship skills	developing a sense of self with others
<b>Item 13</b>	Engages in positive interactions with familiar adults. (Ex: using respectful language or greetings)	Relationship skills	developing a sense of self with others
<b>Item 14</b>	Develops friendships with other children.	Relationship skills	developing a sense of self with others
<b>Item 8</b>	Uses different strategies to solve own problem	Responsible Decision- making	developing a sense of self with others
<b>Item 15</b>	Uses basic strategies for handling typical conflicts. (Ex: sharing, taking turns, and compromising)	Responsible Decision- making	developing a sense of self with others

## 4.2 Model Assumption Test

### 4.2.1 Unidimensionality of TFF instrument

The unidimensionality of the TFF instrument was assessed by principal component analysis of the residuals (PCAR). The most important purpose of unidimensionality is to make sure the instrument meets the assumption of the Rasch model. Multidimensionality can lower the Rasch model validation results accurately. The unidimensionality test result can also provide evidence to determine whether or not all the items contribute to measuring one latent variable, which is an important factor that can influence the assessment quality of the instrument.

The unidimensionality test result below (Table 4.2) shows that there are 63% (25.57) of total raw variance (40.57) explained by the Rasch model, in which the persons explained 39.1% (15.86) and items explained 23.9% (9.71). Overall, the variance explained by the model is more than 50%, which is a good sign for the instrument.

The 1<sup>st</sup> contrast (potential secondary dimension) explained 8% of the unexplained variance (15). The eigenvalue of the 1<sup>st</sup> contrast is 3.26 (including more than three items), which is larger than the cutoff standard value of 3, and it suggests that there may be potential secondary dimensions.

The content validation results suggested that all items' content followed the CASEL and ESD standards, which all fall into the social and emotional competency framework. Additionally, besides the second dimension that exists, the multifaceted construct can create this result also; therefore, further item investigation is needed to decide on adjusting the instrument or not based on the latent variable theoretical framework. After individual

item investigation was processed in the following step, items with misfits have been removed, then the unidimensionality test result has been revised in 4.7.2.

Table 4.2: Standardized Residual Variance

	Eigenvalue	Observed	Expected	
Total Raw Variance in Observation	40.57	100%	100%	
Raw variance explained by measures	25.57	63%	62.8%	
Raw variance explained by persons	15.86	39.1%	39%	
Raw variance explained by items	9.71	23.9%	23.9%	
Raw unexplained variance (total)	15.00	37%	100%	37.2%
Unexplained variance in 1st contrast	3.26	8%	21.7%	
Unexplained variance in 2nd contrast	1.80	4.4%	12%	
Unexplained variance in 3rd contrast	1.46	3.6%	9.7%	
Unexplained variance in 4th contrast	1.21	3.0%	8.1%	
Unexplained variance in 5th contrast	1.14	2.8%	7.6%	

### 4.3 Overall Instrument Quality

The reliability indices of the instrument have been examined to test the reliability of the overall instrument. The result showed the person reliability was .93 and item reliability was .99. For the person, the reliability statistic confirms the children who got higher measure scores actually have a higher-level ability in social and emotional skills compared with children with lower measure scores. For items, the reliability statistic confirmation that the item which has a higher estimate measure score is actually harder than the item with a lower estimate measure score. Both person reliability and item reliability are high, which is all greater than .90. It provided that the item difficulty range is appropriate, and the sample size is good enough to assess this instrument.

The separation indices of the instrument have been examined to understand how well this measure can separate examinees. The person separation is 3.61, and it is greater than 2, which indicates that the instrument is effective to distinguish children based on their social and emotional ability levels. The item separation is 12.17, greater than 2, which implies that the sample size is sufficient to build the item difficulty hierarchy.

Table 4.3: Reliability and Separation of Instrument

Indices	Person	Item
Reliability	.93	.99
Separation	3.61	12.17

#### 4.4 Individual Item Quality Investigation

The Rasch validation can help the researcher to investigate each item's quality by using infit and outfit indexes. Through investigating each item in the instrument, the question about how each item contributes to the social and emotional construct measurement can be answered.

##### 4.4.1 Point- Measure Correlation

The point-Measure correlation index answers the question about the relationship between a person's ability and the person's response to an item. Noticeably positive correlations are expected. A high correlation indicates the person with higher ability should have a higher rating on items. In this instrument, all items showed a noticeable positive correlation from .56 to .81. Items can reveal children's ability in an expected way. The result showed in the Table 4.4.

#### 4.4.2 Item Fit

Examination of each item can provide evidence that all items in this instrument fit the Rasch model. Items with misfits can influence the assessment quality of the measurement. Table 4.4 below shows fit indexes of 15 items of the TFF instrument. Item 6 has misfit in both Infit MNSQ (1.55) and outfit MNSQ (1.6). The value of item 6 is out of the interval from .6 to 1.4. An item with an outfit greater than 1.4 indicates this item is unproductive for construction of measurement. To increase the measuring accuracy of this instrument, item 6 has been removed. The removing process in detail is discussed in 4.7.1.

Table 4.4: Item measure and Fit assessment

Item	Measure	PTmeaure CORP	Infit		Outfit	
			Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
Item 1	1.17	.78	.83	-2.45	.84	-2.09
Item 2	.40	.78	.85	-2.06	.84	-2.10
Item 3	.52	.72	1.06	.78	1.06	.71
Item 4	1.39	.76	.95	-.63	.95	-.55
Item 5	.48	.81	.73	-3.79	.71	-4.04
<b>Item 6</b>	<b>.19</b>	<b>.56</b>	<b>1.55</b>	<b>6.13</b>	<b>1.60</b>	<b>6.48</b>
Item 7	1.09	.78	.95	-.66	.92	-.91
Item 8	1.22	.81	.75	-3.60	.76	-3.41
Item 9	.91	.75	.94	-.85	.96	-.43
Item 10	-1.81	.65	1.31	3.64	1.25	2.67
Item 11	-1.72	.64	1.24	2.89	1.22	2.60
Item 12	-1.28	.71	1.04	.54	1.03	.42
Item 13	-1.40	.73	.93	-.83	.94	-.74
Item 14	-1.15	.73	.95	-.67	1.00	.05
Item 15	-.02	.76	.88	-1.59	.90	-1.24



## 4.5 Differential Item Functioning (DIF)

Differential Item Functioning (DIF) diagnosis provides information if there is any item that has a bias in favor of a particular subgroup but not for other subgroups. In this study, there are two independent variables that can separate the whole sample into more than two groups for the DIF analysis: gender (girls and boys) and site (Churchland, Mt. Hermon, and Olive Branch). In the variable “gender”, there are two subgroups: F (girls, 149) and M (boys, 187). In the variable “site”, there are three subgroups: Churchland (85), Hermon (81), and Olive Branch (129). The differences of each item in difficulty were compared between these subgroups.

### 4.5.1 DIF by Genders

DIF indices were generated by the Winsteps. Rasch DIF contrast, Rasch-Welch *t*-test, and Mantel DIF chi-square indices will be used to test the DIF. All of the items' DIF Contrast are small, ranging from  $-.27$  logit to  $.21$  logit, which is smaller than  $\pm .5$  logit. Additionally, the *t*-test between gender subgroups for all items is not significant  $p > .05$ , which indicates the difference is probably due to randomness error. Additionally, the Mantel chi-square test result showed the difference in gender subgroups is not significant  $p > .5$  for all of the items also. The result showed items did not have a bias for different gender groups.

Table 4.5: Pairwise DIF Analysis of Gender:

		DIF Contrast	Rasch-Welch			Mantel Chi-square	Prob	Item
			T-test	df	Prob			
F	M	-.13	-.75	312	.46	.06	.80	Item 1
F	M	-.06	-.32	313	.75	.15	.70	Item 2
F	M	-0.6	-.31	308	.75	.04	.84	Item 3
F	M	.16	.94	317	.35	.57	.45	Item 4
F	M	-.09	-.52	308	.60	.12	.72	Item 5
F	M	.20	1.15	313	.25	.11	.74	Item 6
F	M	.15	.93	315	.36	.662	.43	Item 7
F	M	-.10	-.58	319	.56	.17	.50	Item 8
F	M	.00	.00	310	1.00	.03	.87	Item 9
F	M	.21	1.18	310	.24	.14	.71	Item 10
F	M	-.27	-1.49	305	.14	5.34	.02	Item 11
F	M	.20	1.12	313	.26	.86	.35	Item 12
F	M	.02	.11	310	.91	1.95	.16	Item 13
F	M	-.17	-.90	307	.37	.23	.63	Item 14
F	M	-.15	-.82	309	.41	1.35	.25	Item 15

The participants in this study are 149 girls and 187 boys. The subtotal of children's ability measure with different gender comparison result showed that the mean measure difference is .91, and the difference is significant  $p < .01$  (Welch  $t = 4.76$ ). The difference is statistical significance, it provides strong evidence that girls have better development in social and emotional behavior compared with boys, as the Figure 4.1 below showed.

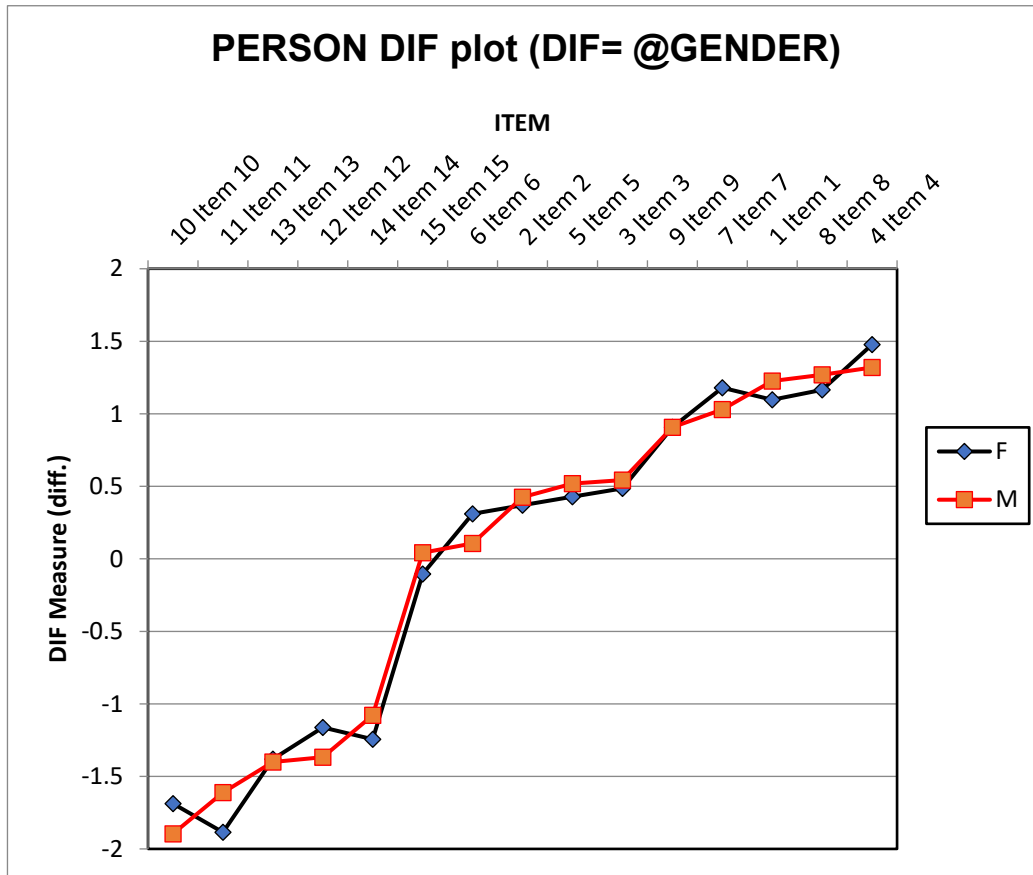


Figure 4.1: DIF Measure Plot of pairwise comparison group by Gender

#### 4.5.2 DIF by Sites

DIF indices were generated by the Winsteps. Rasch DIF contrast, Rasch-Welch *t*-test, and Mantel DIF chi-square indices have been adopted to test the DIF within three different sites. The result showed that item 3, item 4, item 5, item 7, item 8, item 9, item 10, item 11, and item 14 have DIF contrast greater than  $\pm 0.5$  logit between groups, and the *t*-test or Mantel chi-square of these items is statistically significant  $p \leq 0.05$ . Linacre (2012b) suggests keeping items with DIF before further investigation, because the DIF studies are

influenced by sample. The author indicates if the sample is not exactly the same next time, then the biased result may not appear again.

Table 4.6: Pairwise DIF Analysis of Sites:

		DIF	Rasch-Welch		Mantel	Prob	Item	
		Contrast	T-test	df	Prob	Chi-square		
1	2	-.18	-.77	184	.44	1.70	.19	Item 1
1	3	-.09	-.44	189	.66	.31	.58	Item 1
2	3	.09	.42	205	.68	.04	.85	Item 1
1	2	.01	.05	185	.96	.01	.93	Item 2
1	3	-.30	-1.42	190	.16	3.65	.06	Item 2
2	3	-.32	-1.51	202	.13	2.20	.14	Item 2
1	2	.06	.26	185	.80	<.001	.99	Item 3
<b>1</b>	<b>3</b>	<b>-.51</b>	<b>-2.30</b>	<b>188</b>	<b>.02</b>	<b>3.71</b>	<b>.05</b>	<b>Item3</b>
<b>2</b>	<b>3</b>	<b>-.57</b>	<b>-2.63</b>	<b>199</b>	<b>.01</b>	<b>9.19</b>	<b>&lt;.001</b>	<b>Item 3</b>
1	2	.28	1.22	185	.23	.18	.67	Item 4
<b>1</b>	<b>3</b>	<b>.69</b>	<b>3.37</b>	<b>187</b>	<b>&lt;.001</b>	<b>8.46</b>	<b>&lt;.001</b>	<b>Item 4</b>
2	3	.42	2.07	200	.04	2.45	.12	Item 4
1	2	-.36	-1.58	184	.12	4.37	.04	Item 5
1	3	.13	.64	193	.53	.04	.84	Item 5
<b>2</b>	<b>3</b>	<b>.50</b>	<b>2.43</b>	<b>210</b>	<b>.02</b>	<b>6.47</b>	<b>.01</b>	<b>Item 5</b>
1	2	.50	2.10	185	.04	1.02	.31	Item 6
1	3	-.31	-1.47	190	.14	<.001	.96	Item 6
2	3	-0.81	-3.83	199	<.001	3.28	.07	Item 6
<b>1</b>	<b>2</b>	<b>.78</b>	<b>3.56</b>	<b>184</b>	<b>&lt;.001</b>	<b>9.86</b>	<b>&lt;.001</b>	<b>Item 7</b>
<b>1</b>	<b>3</b>	<b>.86</b>	<b>4.33</b>	<b>187</b>	<b>&lt;.001</b>	<b>18.61</b>	<b>&lt;.001</b>	<b>Item 7</b>
2	3	.09	.46	203	.65	.19	.66	Item 7
<b>1</b>	<b>2</b>	<b>-1.00</b>	<b>-4.18</b>	<b>185</b>	<b>&lt;.001</b>	<b>14.23</b>	<b>&lt;.001</b>	<b>Item 8</b>
1	3	-.29	-1.37	190	.17	3.95	.05	Item 8
<b>2</b>	<b>3</b>	<b>.71</b>	<b>3.31</b>	<b>196</b>	<b>&lt;.001</b>	<b>5.66</b>	<b>.02</b>	<b>Item 8</b>
<b>1</b>	<b>2</b>	<b>-.58</b>	<b>-2.51</b>	<b>184</b>	<b>.01</b>	<b>5.73</b>	<b>.02</b>	<b>Item 9</b>
<b>1</b>	<b>3</b>	<b>-.71</b>	<b>-3.35</b>	<b>186</b>	<b>&lt;.001</b>	<b>17.28</b>	<b>&lt;.001</b>	<b>Item 9</b>
2	3	-.13	-0.65	204	.52	1.65	.20	Item 9
1	2	.51	2.23	185	.03	3.06	.08	Item 10
<b>1</b>	<b>3</b>	<b>.73</b>	<b>3.44</b>	<b>202</b>	<b>&lt;.001</b>	<b>10.82</b>	<b>&lt;.001</b>	<b>Item 10</b>
2	3	.22	1.02	208	.31	2.34	.13	Item 10

Table 4.7: Pairwise DIF Analysis of Sites (continued):

1	2	.23	0.94	185	.35	3.34	.07	Item 11
1	3	-.31	-1.43	193	.15	.31	.58	Item 11
<b>2</b>	<b>3</b>	<b>-.54</b>	<b>-2.48</b>	<b>201</b>	<b>.01</b>	<b>2.33</b>	<b>.13</b>	<b>Item 11</b>
1	2	.16	0.65	185	.52	.94	.33	Item 12
1	3	.14	0.64	194	.53	.16	.69	Item 12
2	3	-.02	-0.08	203	.94	.37	.54	Item 12
1	2	.07	0.29	185	.77	.64	.42	Item 13
1	3	-.32	-1.41	196	.16	1.27	.26	Item 13
2	3	-.40	-1.73	202	.09	2.64	.10	Item 13
<b>1</b>	<b>2</b>	<b>-.76</b>	<b>-3.19</b>	<b>184</b>	<b>&lt;.001</b>	<b>6.84</b>	<b>.01</b>	<b>Item 14</b>
1	3	-.17	-.78	194	.44	.95	.33	Item 14
<b>2</b>	<b>3</b>	<b>.59</b>	<b>2.74</b>	<b>211</b>	<b>.01</b>	<b>12.62</b>	<b>&lt;.001</b>	<b>Item 14</b>
1	2	.15	.62	185	.53	.94	0.33	Item 15
1	3	.17	.79	194	.43	.76	0.38	Item 15
2	3	.02	.11	205	.91	.71	0.40	Item 15

In the sample, children who accepted the social and emotional training from their teachers were from three sites: Churchland (85), Hermon (81), and Olive Branch (129). The subtotal of children's ability measure from three different sites comparison result showed, site 1 (Churchland) and site 3(Olive Branch) have a mean measure difference of -.50, which is significant  $p < .05$  (Welch  $t = -2.12$ ). This also indicates children from Olive Branch (site 3) have relatively better development in social and emotional behavior compared with children from Churchland (site 1). Site 2 (Hermon) and site 3(Olive Branch) have a mean measure difference of -.53, the difference is significant  $p < .05$  (Welch  $t = -2.39$ ), which indicates children from Olive Branch have relatively better development in social and emotional behavior compared with children from Hermon. Site 1 (Churchland) and site 2 (Hermon) have a mean measure difference of .02, and the difference is not significant  $p > .05$  (welch  $t = .08$ ), which showed children from Churchland and Hermon

have similar development in social and emotional competency. Figure 4.2 below shows the pairwise comparison between three sites.

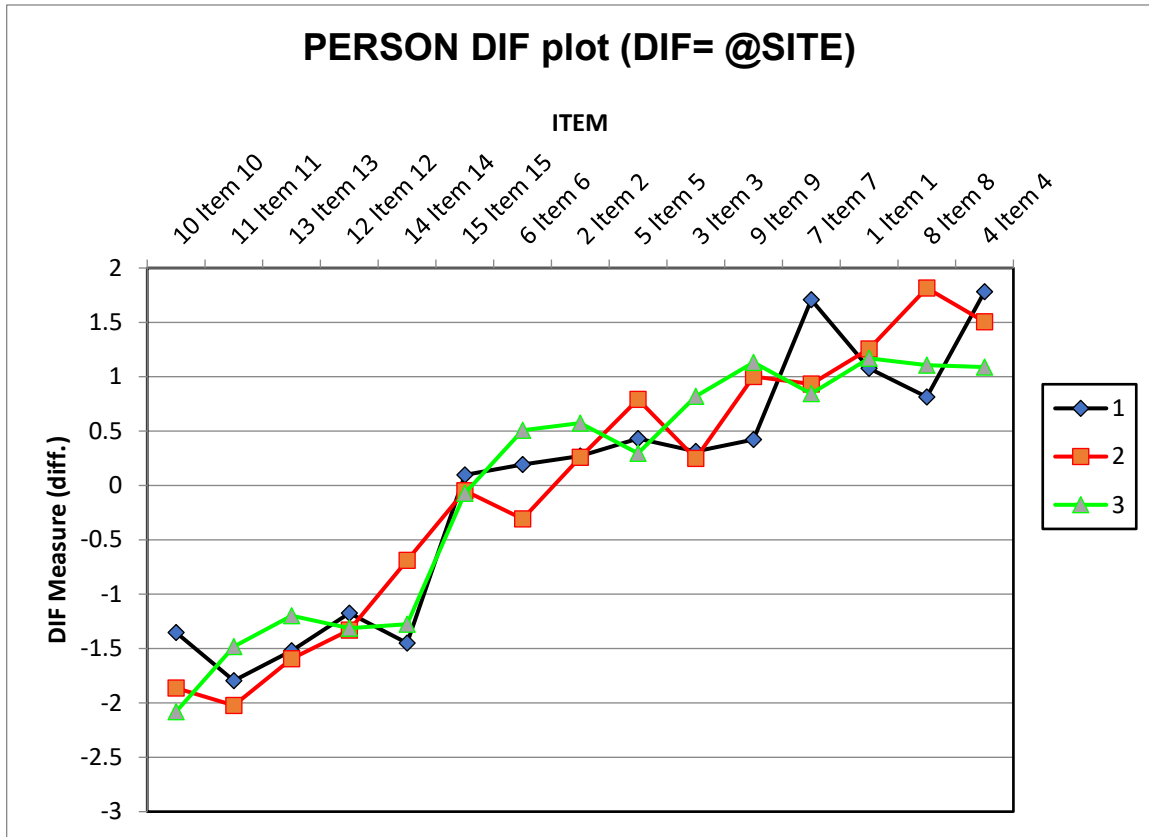


Figure 4.2: DIF Measure Plot of pairwise comparison group by Sites

#### 4.6 Category Andrich Diagnostics

The TFF instrument adopted a 5-point scale design (Never, Rarely, Sometimes, Often, and Always), which is a four-step item design. For accurate measurement purposes, the category order should be in accordance with the latent variable. If disordering is observed, that indicates some of the category's responses are too rarely observed. For each pair of adjacent categories, the distance between the boundaries should fall into the range

of 1.4 logits to 5 logits (J. M. Linacre, 1999). Only Andrich Threshold with disorders was reported in this study.

Table 4.7 showed diagnostic thresholds which revealed the first category (Never) may have disorder risk. There are four items (item 1, item 3, item9, and item 13), and the distance between boundary 1-2 and boundary 2-3 are narrow, smaller than 1.4. If an instrument has a category corresponds to a narrow interval of the latent variable that can create a category disordering issue. Additionally, those five items (item 11, item 12, item 13, item 14, and item 15) that have fewer number of observations (fewer than 10) in category 1, and those eight items (item 1, item 2, item 3, item 4, and item 5, item 7, item 8, item 9) that have fewer number of observations (fewer than 10) in category 5. Figure 4.3 presents the category probability for item 13, which concealed the disordering problem in categories 1 and 2.

Table 4.8: Disorder Andrich Threshold

Item	Boundary 1-2	Boundary 2-3	Distance between adj	Category (Obs. Count <10)
<b>Item 1</b>	<b>-3.24</b>	<b>-1.98</b>	<b>1.26</b>	<b>5</b>
Item 2	-3.29	-1.61	1.68	5
<b>Item 3</b>	<b>-3.18</b>	<b>-2.43</b>	<b>0.75</b>	<b>5</b>
Item 4	-3.43	-1.33	2.1	5
Item 5	-3.34	-1.74	1.6	5
Item 6	-3.35	-1.78	1.57	null
Item 7	-2.89	-1.41	1.48	5
Item 8	-3.84	-1.21	2.63	5
<b>Item 9</b>	<b>-3.27</b>	<b>-2.01</b>	<b>1.26</b>	<b>5</b>
Item 10	-2.86	-1.30	1.56	1
Item 11	-3.32	-1.23	2.09	1
Item 12	-3.60	-1.39	2.21	1
<b>Item 13</b>	<b>-2.43</b>	<b>-2.33</b>	<b>0.1</b>	<b>1</b>
Item 14	-3.23	-1.50	1.73	1
Item 15	-3.21	-1.78	1.43	null

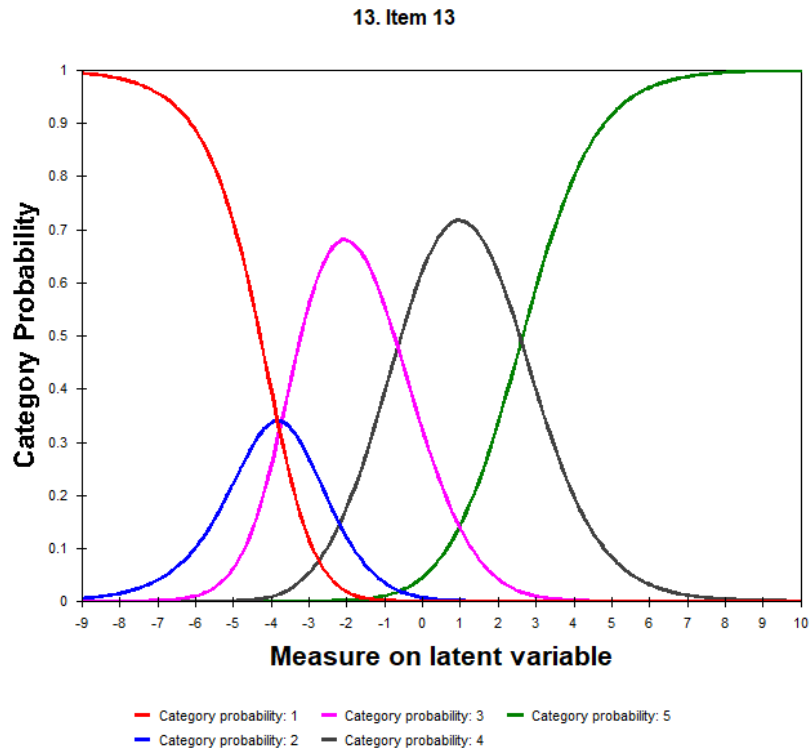


Figure 4.3: measure relative to item difficulty

Linacre (2012a) believes that even the disordered Andrich thresholds are not a statistical problem, which does not violate the Rasch model also. However, disordered categories can influence the measurement practice quality, thus a wide enough gap between Andrich thresholds is needed by researchers. Combining categories that have low observations is a way to solve this problem. Reducing the number of categories can benefit improving the fit of the data to the model also.



## 4.7 The Feeling Friends Instrument Revised

### 4.7.1 Misfit Item Remove Procedure

In the single item investigation step (4.4.2), item 6 showed misfit in both Infit MNSQ (1.55) and outfit MNSQ (1.6). Infit and outfit indices greater than 1.4 indicates unproductive for construction of measurement. Thus, item 6 was removed. Then ran the whole process again. After removing item 6, the remaining items showed good fit (Infit MNSQ and Outfit MNSQ).

Table 4.9: Iterative Procedures to Access of Item Fit

Item removing	Eigenvalue of 1st Contrast	Item Measure	Infit MNSQ	Outfit MNSQ
Item 6	3.24	No item misfit		

In this step, items were investigated independently. The item with misfit (item 6) was removed for improving the overall instrument measure accurately. After this fit diagnosis, there were 14 items left, all reminding items have a good fit.

### 4.7.2 Revised Unidimensionality Test Result

After item 6 was removed from the scale, the revised unidimensionality result is presented below in table 4.9. There was 65.3% (26.30) of total raw variance (40.30) explained by the Rasch model, which shows an improvement compared with the original unidimensionality test result (63%), which the persons explained 45.4% (18.30) and items explained 19.9% (8.01) of variance. Overall, the variance explained by the model is more

than 50%, which is a good sign of the revised instrument. Table 4.9 shows the dimensionality diagnosis result after item 6 was removed.

Table 4.10: Revised Unidimensionality Result

	Eigenvalue	Observed	Expected
Total Raw Variance in Observation	40.30	100%	100%
Raw variance explained by measures	26.30	65.3%	65.1%
Raw variance explained by persons	17.74	44%	43.9%
Raw variance explained by items	8.59	21.2%	21.2%
Raw unexplained variance (total)	14.00	34.7%	100% 34.9%
Unexplained variance in 1st contrast	3.24	8%	23.2%
Unexplained variance in 2nd contrast	1.58	4.3%	11.3%
Unexplained variance in 3rd contrast	1.47	3.6%	10.5%
Unexplained variance in 4th contrast	1.18	2.9%	8.5%
Unexplained variance in 5th contrast	1.03	2.5%	7.3%

The 1<sup>st</sup> contrast (potential secondary dimension) explained 8% of the unexplained variance (14). The eigenvalue of the 1<sup>st</sup> contrast is 3.24 (including nearly three items), which is greater than the cutoff standard value of 3. To further investigate the unidimensionality of this instrument, the next step is to review clusters in the first contrast by items' loading.

Table 4.11: First Contrast Loading and Cluster Group

	Item	Loading	Cluster
A	Item 4	.68	1
B	Item 3	.53	1
C	Item 8	.50	1
D	Item 1	.47	1
E	Item 2	.35	2
F	Item 7	.32	2
G	Item 9	.20	2
g	Item 5	.17	2
a	Item 13	-.65	3
b	Item 10	-.63	3
c	Item 11	-.60	3
d	Item 12	-.57	3
e	Item 14	-.43	3
f	Item 15	-.24	3

After removing item 6, the rest of items loading is shown in table 4.10. Eight items (4, 3, 8, 1, 2, 7, 9) had positive loading and eight items (13, 10, 11, 12, 14, 15) showed negative loading. Positive loading items mostly focus on measuring children's sense of self and others. While negative loading items tend to measure children's social skills and interpersonal relationships.

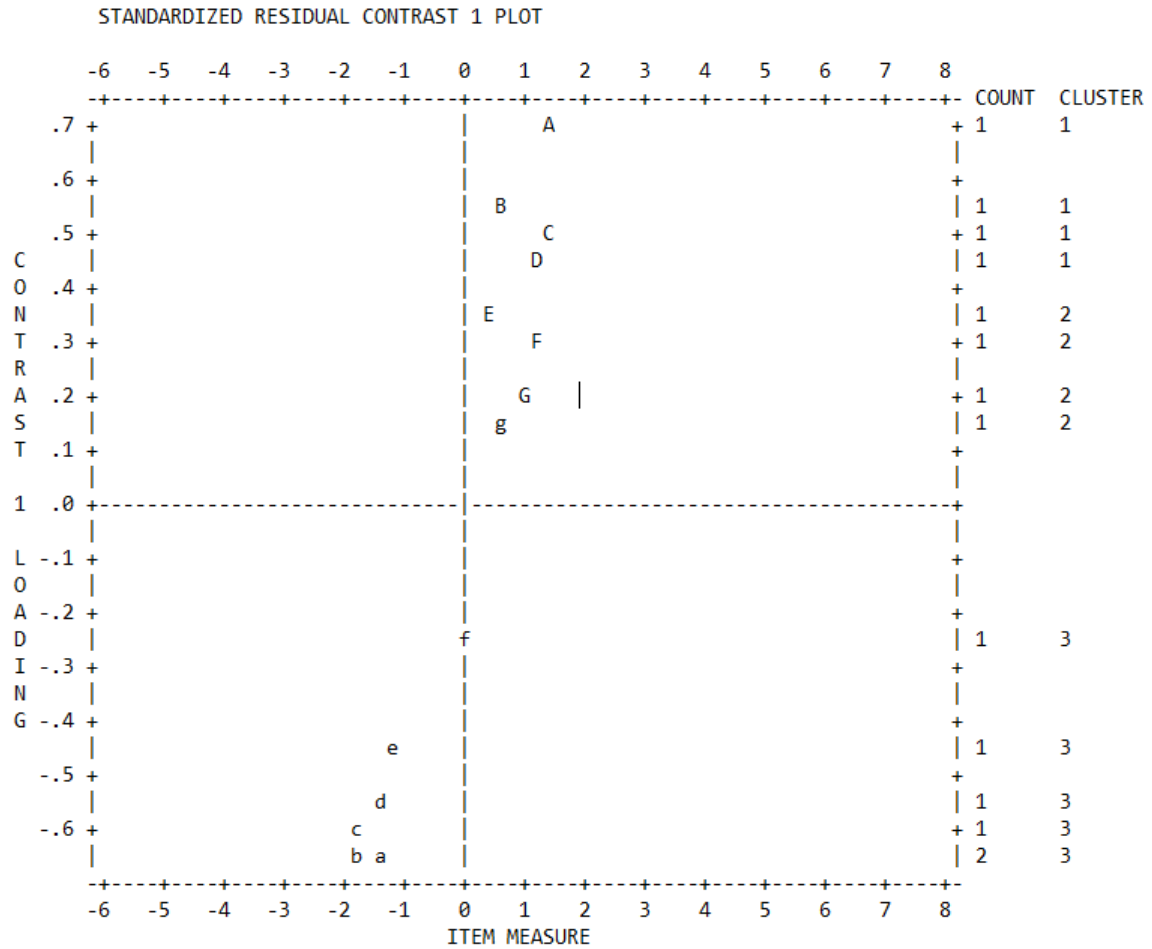


Figure 4.4: Revised First Contrast Standardized Residual Plot

The Pearson's correlation coefficients of person measure in the first contrast between clusters are reviewed next. High disattenuated correlation indicates that person measures on the item clusters compared are sharing a majority of variance, probably part of the same dimension. On the opposite, a low disattenuated correlation suggests that two clusters are measuring different contrasts. Table 4.11 presents the result of the correlation coefficient of item clusters within the first contrast. Cluster 1 and cluster 2 were highly correlated (1.00) and so were cluster 2 and 3 that indicate (.83). Cluster 1 and cluster 3 are moderated correlated (.68), which indicates that person measures on the two item clusters

have more than half as much variance in common as they are independent. Three pairs of clusters correlation are all greater than .57, which suggests that all items are measuring only one latent variable (J. M. Linacre, 2012a).

Table 4.12: Correlation Coefficient of Item Clusters within First Contrast

PCA Contrast	Item Clusters	Pearson Correlation	Diattenuated Correlation
1	1-3	.59	.68
1	1-2	.86	1.00
1	2-3	.71	.83

#### 4.7.3 Revised Item Measure and Fit

An examination of each item fit has been implemented in the revised instrument. Table 4.12 showed the 14 items fit the original TFF instrument. The result showed no misfit item in the revised instrument. All of the items' Infit MNSQ and outfit MNSQ fall within the acceptable interval, from .6 to 1.4.

Table 4.13: Revised Item Measure and Fit

Item	Measure	PTmeasure		Infit		Outfit	
		CORP	MNSQ	ZSTD	MNSQ	ZSTD	
Item 1	1.24	.78	0.88	-1.64	0.89	-1.38	
Item 2	0.43	.78	0.90	-1.36	0.89	-1.44	
Item 3	0.55	.71	1.13	1.56	1.14	1.56	
Item 4	1.48	.77	0.98	-0.30	0.97	-0.30	
Item 5	0.51	.82	0.76	-3.34	0.73	-3.62	
Item 7	1.16	.77	1.02	0.23	0.99	-0.08	
Item 8	1.3	.81	0.76	-3.42	0.77	-3.18	
Item 9	0.97	.76	0.95	-0.62	0.97	-0.29	
Item 10	-1.88	.66	1.33	3.90	1.27	2.71	
Item 11	-1.78	.64	1.27	3.27	1.27	2.82	
Item 12	-1.32	.71	1.09	1.22	1.09	1.08	
Item 13	-1.46	.72	1.00	-0.02	1.00	0.08	
Item 14	-1.19	.72	1.02	0.29	1.28	3.11	
Item 15	-0.01	.78	0.87	-1.66	0.89	-1.34	

#### 4.7.4 Wright Map

In the Wright Map, the left side represents children's measures, and the right side represents items to measure. The means of person ability and item difficulties of the instrument are nearly (-.01 and 0 respectively). On the person side, the higher location indicates higher ability; On the item side, the higher location means more difficulty. The item on the higher location has higher endorsed difficulty than item in the lower position. For example, in this map, the item 4 is the hardest item and the item 10 is the easiest item in the scale.

The means of person measure and item measure were nearly the same (-.01 and 0 respectively). Children's ability ranges from -5.75 logit to 7.59 logit; items difficulty ranges from -1.88 to 1.48; children's social and emotional competency range is wider than item

difficulty range, which suggests this instrument is appropriate for measuring children with moderated social and emotional ability. More challenge items and basic items can be added to better measure children's extraordinary and relatively slower development in social and emotional competency.

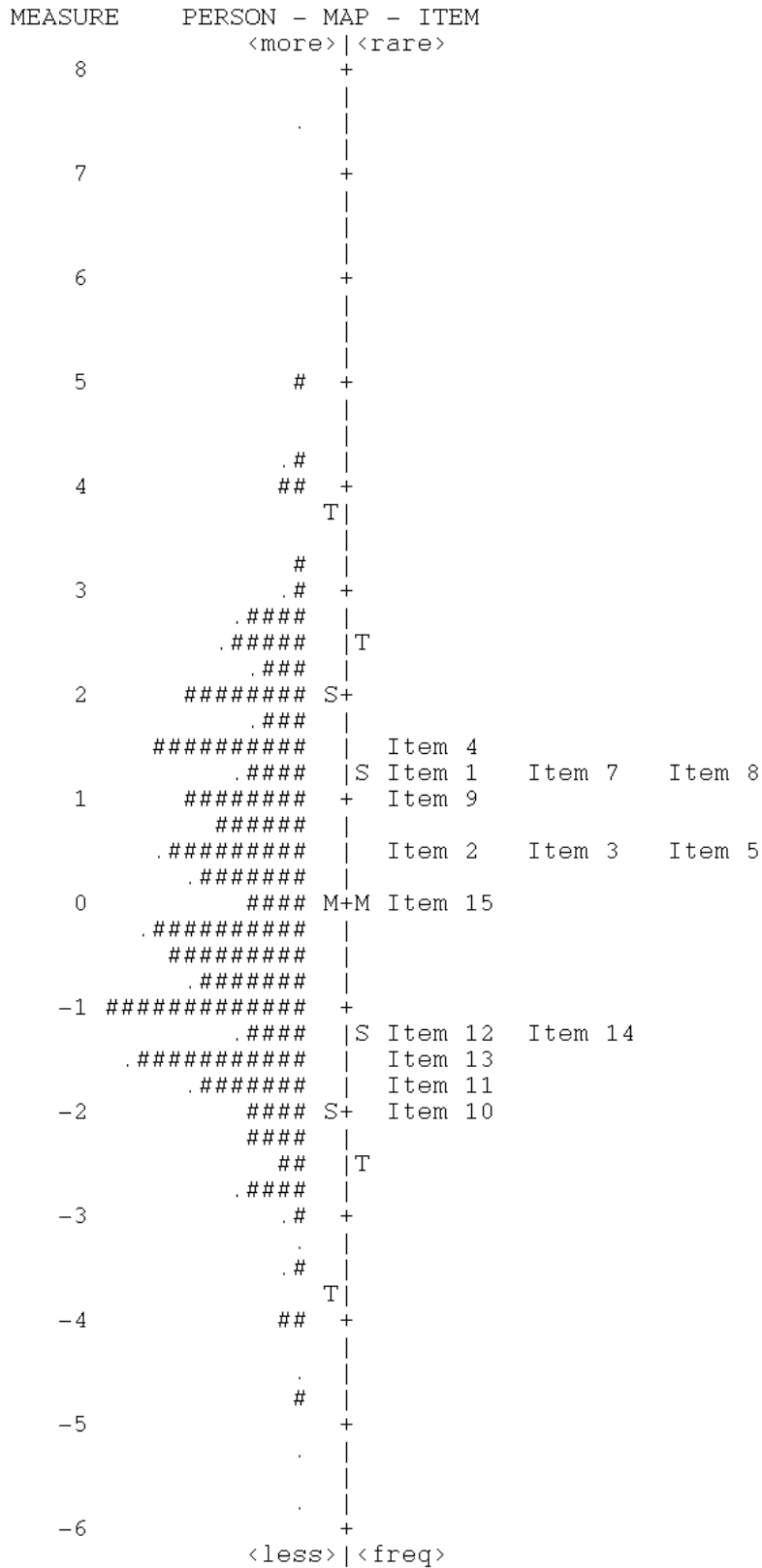


Figure 4.5: Wright Map



#### 4.7.5 Keyform

The keyform presented each item and their response categories and item difficulty order (Figure 4.6). The left side is the predicted average response for each category. For instance, for item 4, when children social and emotional ability between -2.38 to -.01, then we expect to observe category “2” was selected. For children’s competency lower than 2.38, then we predict they select “1” in item 4. For partial credit model analysis, the KeyForm can display category structure for each item, which can recognize each item category step pattern.

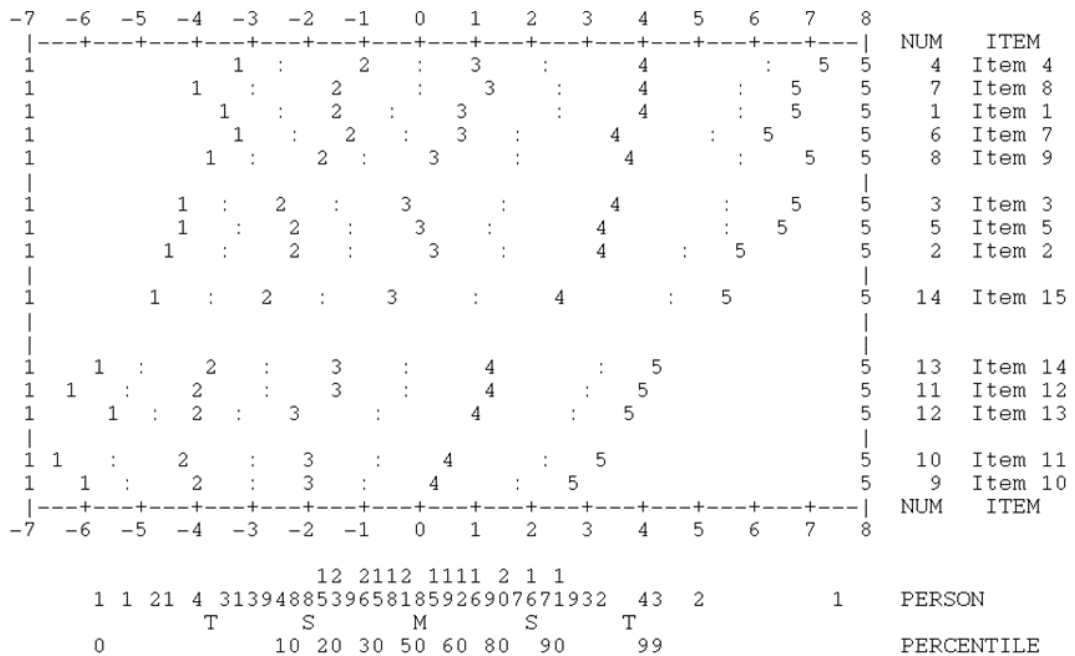


Figure 4.6: Keyform

#### 4.8 Pre and Post Assessments Result Comparison

The stacking method is adopted to assess the children’s social and emotional ability change after training. With this method, the Rasch model can be used to transfer the

original ordinal score to continuing measure number and calibrate item difficulty estimate, children's pre-ability estimate, and children's post ability to calibrate in the same ruler for comparison. In this study, 336 children had been assessed in the pre-assessment, and 289 children had been assessed in the post-assessment. The children's sample groups in pre- and post-assessment are homogeneous.

The pre- and post-data were combined in one dataset. A column named "Time" was added to represent the measured result from pre and post. In the *t*-test, the Time will be used as an independent variable to test the change over time. Item 6 was removed from the data due to the item misfit discussed before. Therefore, in the stacking comparison, 14 items of assessment results included in, the original score range of the revised instrument is from 14 to 70, the higher, the better.

#### 4.8.1 Stacking Result

The stacking result plot is shown in Figure 4.7. In the plot, the orange line is the regression line showing the best fit to the data; the black dash line represents the identity line of perfect alignment. In the figure, we can clearly tell most children were plotted above the perfect alignment line, which indicates most of children have better performance in the post-assessment. That means the training has positive influence in most children's social and emotional ability development.

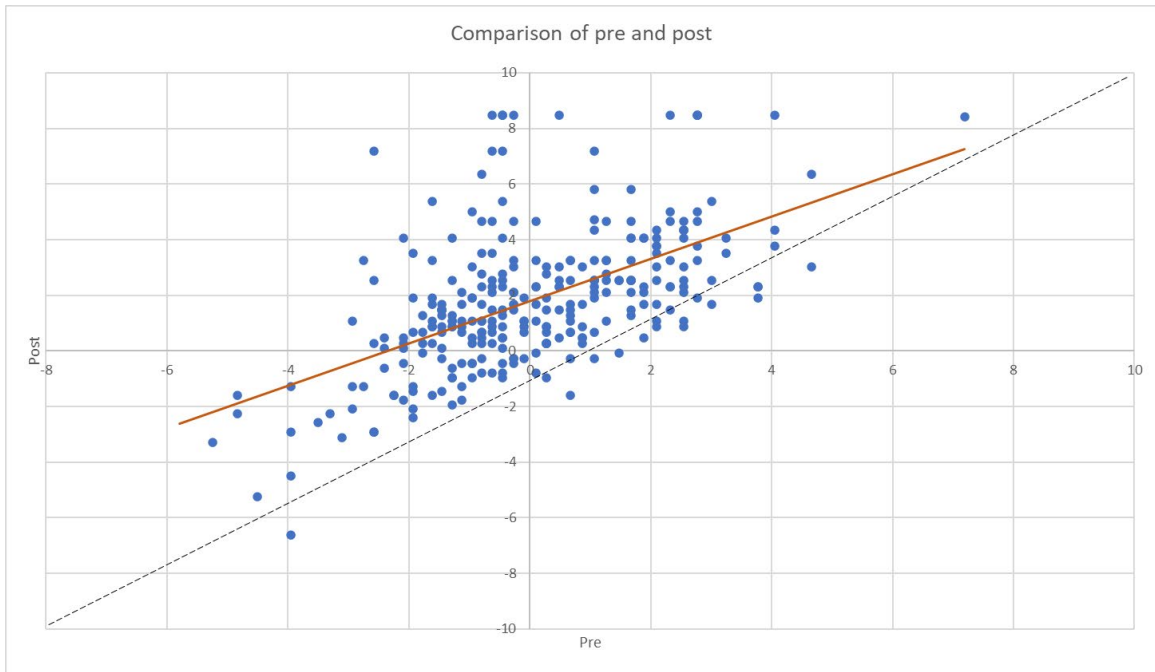


Figure 4.7: Stacking Comparison Plot

#### 4.8.2 T-test

For a larger number of samples, which cannot compare location change one by one, a *t*-test can be applied to test the change over time points and to test the statistical significance of the change.

In this study, the null hypothesis is that there is no significant difference in children’s social and emotional competency before and after the training is received. The  $\alpha$  level was set at .05. The JMP 16 Pro was adopted to apply the *t*-test on the pre-assessment and post-assessment results. The Rasch measurement score was used as the dependent variable, and the time (pre- and post-) was used as the independent variable. The comparison and *t*-test results are shown in Table 4.13 below.

Table 4.14: pre- and post- t-test result

	<b>Pre</b>	<b>Post</b>
<b>N</b>	336	289
<b>Min</b>	-5.79	-6.62
<b>Max</b>	7.19	8.49
<b>Mean</b>	0.07	1.81
<b>Std Dev</b>	1.86	2.50
<b>Lower 95%</b>	-.13	1.52
<b>Upper 95%</b>	.27	2.10
<b>Difference</b>		1.74
<b>Lower CL Dif</b>		1.39
<b>Upper CL Dif</b>		2.10
<b>Confidence</b>		0.95
<b>T- Ratio</b>		9.76
<b>DF</b>		526
<b>Probability&gt;  t </b>		<.001

The mean difference between pre- and post-assessment is 1.74 logit, which was shown to be statistically significant by the *t*-test ( $p < .05$ ,  $t=9.76$ ). The result provided strong evidence to reject the null hypothesis, which suggested that children have significant improvement after accepting the social and emotional training.

#### 4.9 Chapter Summary

In this chapter, the diagnoses of the instrument validation and pre- and the post-comparison results are presented. This instrument design followed the content of CASEL and ESD standards. Principal Component Analysis Residual diagnoses showed that all of

the items contributed to measuring a single construct. The overall instrument has good reliability and validity. One item (item 6) has a misfit, which may create a disregard problem, which can be removed to keep measurement accurate. The DIF analysis showed that there was no item DIF found by gender. However, girls showed better development in social and emotional competency compared with boys. Items 3, item 4, item 6, item 7, item 8, item 9, item 10, and item 11 showed DIF by sites. As Linacre (2012b) suggests that items with DIF were kept in this study because DIF was sample dependent, which means that if the sample was not exactly the same the next time, then the biases might not appear again. Site subgroups' results showed children from site 3 (Olive Branch) had better development in social and emotional competency compared with site 1 and site 2. Children from site 1 and site 2 had similar developments in social and emotional competency.

In the category design diagnosis, the result showed a disordering issue in a couple of items. That may be because there were not enough observations in a particular category, or the meaning of the category was not clear. A combination between categories one and two can be considered to solve this problem because a couple of items showed not enough distance between boundaries one-two and two-three. Additionally, category 1 has less than 10 observations for a couple of items. After combining two categories, the instrument will become a four-point rating scale instrument. Researchers believe a scale without a middle point can provide a better measurement (K. D. Bradley et al., 2015a). However, in keeping with the data information completeness, category collapse was not adopted in this study.

Instrument and data are modified by the diagnosed instrument before applying the stacking comparison. Item 6 and related responses were removed from both pre- and post-assessment results. A t-test was adopted to understand how children has change over time.

The *t*-test result showed that after the training, children make a statistically significant improvement in social and emotional competency.

## CHAPTER 5. CONCLUSION AND DISCUSSION

For a long time, social and emotional education has been placed in a secondary position, behind cognitive education (for example, math and reading). In October 2015, the release of the Every Student Succeeds Act emphasized the importance of preschool education, and it promoted a well-rounded education system. Consequently, social and emotional education has been included as an important component and indicator of the education system.

Preschool education is the first stage of education, generally for children from 3-5 years old. In this period, the education goal is to prepare children to be ready for kindergarten and further academic education. School readiness includes cognitive, behavioral, and social-emotional readiness. Specifically, social-emotional readiness involves self-confidence, the development of a positive relationship, the ability to listen, and the skill to solve social problems. Multiple groups take accountability for social and emotional development of children. Schools also should take accountability to create a safe, healthy, and supportive school climate for children. Children's personalities also can influence the development of social and emotional ability.

The social-emotional ability has a positive influence on children to develop positive relationships, which includes building positive interactions with peers and even with adults such as teachers and parents. Social-Emotional ability also makes a positive influence on students' academic performance. Researchers claimed that students with higher social-emotional ability are more easily successful in their academic work. And in the long run, people with high social-emotional ability tend to achieve quality of life and happiness.

Recently, schools and parents have been gradually emphasizing social and emotional development as an important part of children's education content. However, because there are limitations for young students to express their ideas fully, it is difficult for both school and parents to measure children's social and emotional development accurately. Currently, most instruments focused on measuring social and emotional ability are originally designed for school-age students, who have better expression ability. Therefore, these instruments cannot provide an accurate result to the preschool children group. As in this case, researchers and educators sometimes need to customize a new measurement tool by themselves.

A new problem raised by using personality instruments is the quality of the instrument. Classic Test Theory (CTT) methods are widely adopted by researchers to assess the instrument performance. However, CTT has some inherent limitations; therefore, an advanced method is required for researchers to assess an instrument's performance. In this study, the Rasch model has been introduced to provide a comprehensive guide to researchers to assess the quality of an instrument that they design, adopt, or organize based on their research requirements and research purpose. Comparing with the CTT theory validation method, Rasch validation can be applied to an instrument with a small number of items and provide diagnosis for both the instrument and each item. In this study, the equating method (stacking) was adopted to compare repeat measurement results (pre- and post-) to assess the social and emotional training program.

Chapter 2 reviewed the definition of social and emotional education, social and emotional education history, related policy and standard, social and emotional learning



accountability, and how social and emotional ability influence a child from short term to lifetime. Additionally, the theory of the Rasch model has been introduced.

Chapter 3 introduces the process of using the Rasch model in detail to assess an instrument to ensure the quality of an instrument by comprehensive and thorough inspection. Chapter 4 follows the steps of chapter 3, and based on the preset standards, it (chapter 4) makes the judgment of the instrument's quality and uses the result to modify the instrument. Then the stacking comparison was applied to assess the training influence on children's social and emotional development. In this chapter, the validity of this instrument will be suggested. Additionally, the limitations and contributions of this study and future research directions are also discussed in this chapter.

## 5.1 Research Question Results

### 5.1.1 Research Question One

(1) How well does the instrument measure the preschool children's social and emotional learning situation?

This question was answered by testing the Rasch unidimensionality assumption, reliability, and separation indices. Additionally, CASEL and NC Emotional and Social Development (ESD) were adopted to review item content, make sure all items fall into the official social and emotional framework and group items, and make sure all items contribute to measuring preschool children's social and emotional competency.

The unidimensionality test of the original instrument result shows that the 1<sup>st</sup> contrast (potential secondary dimension) eigenvalue is 3.26 (including more than three items), which is larger than the cutoff standard value of 3, and it suggests that there may

be potential secondary dimension that exists. After removing the misfit item (item 6), the eigenvalue of 1<sup>st</sup> contrast is 3.24, which is still nearly 3 items. Item loading and the correlation coefficients of person measure in the first contrast between clusters of revised instruments are reviewed. Results suggested that all items were measuring only one latent variable.

From the content validity result, the items all fall in the range of CASEL and NC's Emotional and Social Development (ESD). The contents are appropriate for measuring preschool children's development in social and emotional skills.

Overall, based on the result of unidimensionality, separation, reliability, and content review, this whole instrument is well designed, and the items are appropriate to the content. This instrument can effectively group children based on their real social-emotional competency and can provide a reliable result in measuring children's social and emotional development.

#### 5.1.2 Research Question Two

(2) How well does each item contribute to the social and emotional construct measurement (how well does each item fit into the Rasch model)?

To solve this problem, the Rasch infit and outfit indices were adopted to test each item's quality. The output showed that item 6 infit and outfit larger than 1.4, which indicates it is unproductive for the construction of measurement. Thus, Item 6 was removed. After that, remaining items (14 items left) in the revised instrument have a good fit. Item 4 (Can describe the impact their behavior has on others) is the hardest item for children to get high rates from teachers. Item 10 (Transitions between activities without getting upset) is the

easiest one for children to get a high-rated score from their teachers. It suggests that for preschool children, controlling their emotions while changing is not very hard while understanding the result of their behaviors toward others is difficult.

A differential item functioning diagnosis was also applied to the revised scale to make sure items were fair to different subgroups. In this study, the gender subgroups (girls and boys) and site subgroups (Churchland, Mt. Hermon, and Olive Branch) were examined. The result showed items do not have a bias for different gender groups. In the subgroups of sites, children from site 3 (Olive Branch) showed better performance on the whole scale. One possible reason is that nine items have item bias designed in favor of children from Olive Branch. Another possibility reason is that the sample group of children in Olive Branch ends to show better development in social and emotional abilities. Without further evidence of item bias, then items showed DIF are suggested to remain.

Wight map (Figure 4.5) showed that most items that fall into the social awareness are in the higher location, indicating that preschool children showed difficulty in this area, which is also a challenge for teachers to deliver this content to preschool children. Social skills items are calibrated in the lower position of the map, which means for children, social practice with others seems easier to learn.

Moreover, some children's ability is higher than the most difficult item in this instrument, so adding more challenging items can improve the accuracy of the instrument in measuring children with extraordinary development in social and emotional competency.

### 5.1.3 Research Question Three

(3) How well does the category design of the social and emotional instrument fit into the Rasch model?

To fit into the Rasch model, the design of scale structure (rating scale design) was examined by the Rasch category threshold. Originally, this instrument adopted a 5-point rating scale, which is a middle point. However, recently researchers claimed that even number designs, which do not have middle points, can provide a more accurate measurement. The result showed that four items contained disorder problems (the distance between boundaries 1-2 and boundary 2-3 is narrow). When a category corresponds to a narrow interval of the latent variable, that can create category disordering. Additionally, 13 out of 15 items have less than 10 observations either in category 1 or category 5.

Linacre (2012a) suggested that even disordered Andrich thresholds are not a statistical problem, which does not violate the Rasch model also. However, the disordered category can influence the measurement practice quality. A wide enough distance between Andrich thresholds is needed to improve measurement accuracy. Combining categories with a small number of observations is a way to solve this problem. Also, reducing the number of categories may improve fitting the data into the model. Combining either category 1 and category 2 or category 4 and category 5 can be considered.

### 5.1.4 Research Question Four

(4) How well does the Stacking Comparison method assess children's social and emotional learning improvement after the training?

To answer this question, the equating method (stacking) was applied to the pre- and post-assessment results, and by comparing the changes between the results the social and emotional learning program could be assessed properly. Additionally, a *t*-test had been used to test the statistical significance of the change between pre- and post-emotional assessment results.

The result suggested that social and emotional training is effective in improving preschool children's social and emotional competency. After 10 weeks of training, on average, children improved by 1.74 logit in comparison with pre-training. The *t*-test result is significant, which provides strong evidence to support children's improvement after accepting the training.

## 5.2 Discussion

### 5.2.1 Stacking Comparison Assessment Tool

Education evaluation currently plays an important role to aid in making an administrative decision. An education program is an organized collection of planned activities designed to reach certain education objectives to solve a certain problem (Royse et al., 2015). Education targets should be determined well for the realization of efficiency. Education intervention programs are expected to have some positive influence on program participants. It is important to know to what extent the education program achieves its educational goals. The most important way to determine this is to evaluate the program. Evaluation is the process of measuring if the output of a program meets its preset target through assessing the input and activities implemented in the project. The aspects that the

programs have contributed to the aim of education can be fully evaluated in the field of education using qualitative, quantitative, or a combination of distinctions (Basaran et al., 2021). A well-designed preschool program tends to produce a long-term positive influence on academic performance (Barnett, 2008). Preschool programs should regularly assess children's learning and development to monitor how well they are accomplishing their goals (Barnett, 2008).

Education results will be an important source for making the decision to start a new program or continue and improve the quality of an existing program (Royse et al., 2015). Evaluation provides the foundation for learning. If we do not "measure" it, it does not count (Cohen, 2006). Rasch model can be a robust tool in measurement design and result in measurement result assessment to provide evidence for evaluation purposes.

Generally, repeatable measurement is adopted by educators and researchers to assess the effectiveness of a program. For comparing the measurement score from multiple time points, an equating method is needed to create a relationship among multiple measurement results. Equating methods (stacking and racking) with Rasch model can provide an advanced solution to make a relationship in measurement results. Rasch model have theoretical advantages compared with the CTT methods. Rasch model has an invariance attribute, which means the item parameters and person parameters are independent. In the theory, if the data fit the assumption of Rasch, then it is possible to estimate the examinees' ability no matter which forms of test the examinees took. The difficulty in estimating the item will not be influenced by the examinee's ability estimate, and the examinees' ability estimation will be constant if the items' difficulty changed. No matter which forms of test the examinees take, the examinees' ability estimate will be

constant. The invariance attribute of Rasch model makes the test bank applied realistic, which can provide researchers with more flexibility in administrating tests in the future.

### 5.3 Contribution

Currently, there is a growing demand for measuring preschool children's social and emotional competency development. However, currently, widely adopted instruments are mostly focused on school-age children. School-aged instruments are not appropriate to apply to preschool children who do not have enough explanatory ability. Additionally, these instruments are designed to cover as large extended content as possible and try to cover all the aspects of the construct. Therefore, these instruments have a very large item number, which creates difficulty to administrate. A well-designed instrument that covers required aspects with a reasonable number of items is always desirable for educators and researchers. For providing reliable measurement results, the instrument validation process is essential.

In this study, the advanced model which can easily be adopted by researchers to assess a new instrument is introduced and provides a comprehensive guide of Rasch model validation diagnoses including reliability, validity, single item fit, DIF, and category threshold. The Rasch model also provides a solution for validating instruments with small items, or for small number samples. Additionally, Rasch model validation is not influenced by missing data. The Rasch validation method provides more possibility for researchers to design, assess, and adopt a new instrument.

## 5.4 Limitations

### 5.4.1 Item Range

This instrument originally contained 15 items. After the misfit removal process, 14 items were left. A small number of items in an instrument can save resources in administration. However, 14 items cannot fully cover all the aspects of preschool social and emotional teaching standards. More items can be added to extend the measurement scope of this instrument.

### 5.4.2 Control Group Design

In this study, all the children who participated in the pre-and post-assessments accepted the training. Therefore, for program result assessment comparison, there does not have control group comparison results. To better assess the program's effectiveness, a control group design is needed. To compare the children who accepted the training with children who did not accept the training, testing if there is a statistically significant difference between these two groups can provide more powerful evidence of the program training result.



APPENDICES

APPENDIX 1. ORIGINAL TFF INSTRUMENT

**The Feeling Friends at School Pre-Assessment**

*Directions:* Rate the child’s social-emotional skills **prior to** the implementation of the Feeling Friends curriculum using the Likert-type rating scale below. Place a check in the box that indicates the child’s performance in each area.

	Never	Rarely	Sometimes	Often	Always
<b>Social and Self-Awareness and Responsible Decision Making</b>					
Makes empathetic statements to other children.					
Can talk about own feelings					
Identifies basic emotions in books or photographs.					
Can describe the impact their behavior has on others.					
Communicates feelings in manners that are appropriate to the situation.					
Engages in some positive interactions with strangers. (ex: parent volunteers)					
Demonstrates an understanding of their own uniqueness. (ex: talents, interests, preferences, or culture)					
Uses different strategies to solve own problems					
<b>Self-Management and Relationship Skills</b>					
Uses various coping techniques to manage emotions with the support of an adult.					
Transitions between activities without getting upset.					
Able to follow classroom rules.					
Seeks help from adults when problems arise. something to do together, joining an existing activity, or sharing a toy)					
Engages in positive interactions with familiar adults. (ex: using respectful language or greetings)					
Develops friendships with other children.					

Uses basic strategies for handling typical conflicts. (ex: sharing, taking turns, and compromising)					
---	--	--	--	--	--

APPENDIX 2. REVISED TFF INSTRUMENT

**The Feeling Friends at School Pre-Assessment**

*Directions:* Rate the child’s social-emotional skills **prior to** the implementation of the Feeling Friends curriculum using the Likert-type rating scale below. Place a check in the box that indicates the child’s performance in each area.

	Never	Rarely	Sometimes	Often	Always
<b>Social and Self-Awareness and Responsible Decision Making</b>					
Makes empathetic statements to other children.					
Can talk about own feelings					
Identifies basic emotions in books or photographs.					
Can describe the impact their behavior has on others.					
Communicates feelings in manners that are appropriate to the situation.					
Demonstrates an understanding of their own uniqueness. (ex: talents, interests, preferences, or culture)					
Uses different strategies to solve own problems					
<b>Self-Management and Relationship Skills</b>					
Uses various coping techniques to manage emotions with the support of an adult.					
Transitions between activities without getting upset.					
Able to follow classroom rules.					
Seeks help from adults when problems arise. something to do together, joining an existing activity, or sharing a toy)					
Engages in positive interactions with familiar adults. (ex: using respectful language or greetings)					
Develops friendships with other children.					
Uses basic strategies for handling typical conflicts. (ex: sharing, taking turns, and compromising)					

## REFERENCES

- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4), 561–573. <https://doi.org/10.1007/BF02293814>
- Anson, Amyr., Cook, T. D., Habib, F., Grady, M. K., Haynes, N., & Comer, J. P. (1991). The Comer School Development Program: A Theoretical Analysis. *Urban Education*, 26(1), 56–82. <https://doi.org/10.1177/0042085991026001005>
- Aos, S., Lieb, R., Mayfield, J., Miller, M., & Pennucci, A. (2004). *Benefits and costs of prevention and early intervention programs for youth*.
- Arnold, D. H., Kupersmidt, J. B., Voegler-Lee, M. E., & Marshall, N. A. (2012). The association between preschool children's social functioning and their emergent academic skills. *Early Childhood Research Quarterly*, 27(3), 376–386.
- Bagdi, A., & Vacca, J. (2005). Supporting early childhood social-emotional well being: The building blocks for early learning and school success. *Early Childhood Education Journal*, 33(3), 145–150.
- Barnett, W. S. (2008). *Preschool education and its lasting effects: Research and policy implications*.
- Basaran, M., Dursun, B., Gur Dortok, H. D., & Yilmaz, G. (2021). Evaluation of Preschool Education Program According to CIPP Model. *Pedagogical Research*, 6(2).
- Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., Owen, M. T., & NICHD Early Child Care Research Network. (2007). Are there long-term effects of early child care? *Child Development*, 78(2), 681–701.
- Berk, L. E. (2014). *Development through the lifespan* (Sixth edition). Pearson.

- Bierman, K. L., & Motamedi, M. (2015). Social and emotional learning programs for preschool children. *Handbook of Social and Emotional Learning: Research and Practice*, 135–151.
- Blewitt, C., O'Connor, A., Morris, H., May, T., Mousa, A., Bergmeier, H., Nolan, A., Jackson, K., Barrett, H., & Skouteris, H. (2019). A systematic review of targeted social and emotional learning interventions in early childhood education and care settings. *Early Child Development and Care*, 0(0), 1–29.  
<https://doi.org/10.1080/03004430.2019.1702037>
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014a). Fit. In W. J. Boone, J. R. Staver, & M. S. Yale (Eds.), *Rasch Analysis in the Human Sciences* (pp. 159–189). Springer Netherlands. [https://doi.org/10.1007/978-94-007-6857-4\\_8](https://doi.org/10.1007/978-94-007-6857-4_8)
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014b). How Well Does That Rating Scale Work? How Do You Know, Too? In W. J. Boone, J. R. Staver, & M. S. Yale (Eds.), *Rasch Analysis in the Human Sciences* (pp. 191–216). Springer Netherlands. [https://doi.org/10.1007/978-94-007-6857-4\\_9](https://doi.org/10.1007/978-94-007-6857-4_9)
- Bowman, B., Donovan, M., & Burns, S. (n.d.). & the Committee on Early Childhood Pedagogy of the National Research Council. (2000). *Eager to learn: Educating our preschoolers*.
- Bracken, B. A., Keith, L. K., & Walker, K. C. (1998). Assessment of preschool behavior and social-emotional functioning: A review of thirteen third-party instruments. *Journal of Psychoeducational Assessment*, 16(2), 153–169.
- Bradley, K. D., Cunningham, J. D., Akers, K. S., & Knutson, N. (2011). *Middle category or survey pitfall—Using Rasch modeling to illustrate the middle category*. 21.

- Bradley, K. D., Peabody, M. R., Akers, K. S., & Knutson, N. (2015). Rating Scales in Survey Research: Using the Rasch model to illustrate the middle category measurement flaw. *Survey Practice*, 8(1), 2852. <https://doi.org/10.29115/SP-2015-0001>
- Bradley, K., & Sampson, S. (2005). A case for using Rasch Rating Scale analysis to assess the quality of measurement in survey research. *The Respondent*, 12–13.
- Brooks-Gunn, J. (1997). *Neighborhood poverty: Context and consequences for children* (Vol. 1). Russell Sage Foundation.
- Brownell, C. A., Burchinal, M., & Caldwell, B. (2003). Does quality of child care affect child outcomes at age 4 1/2. *Developmental Psychology*, 39(3), 451–469.
- Budescu, D. (1985). Efficiency of Linear Equating as a Function of the Length of the Anchor Test. *Journal of Educational Measurement*, 22(1), 13–20.  
<https://doi.org/10.1111/j.1745-3984.1985.tb01045.x>
- Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record*, 112(3), 579–620.
- Cannon, J. S., Kilburn, M. R., Karoly, L. A., Mattox, T., Muchow, A. N., & Buenaventura, M. (2017). *Investing Early: Taking Stock of Outcomes and Economic Returns from Early Childhood Programs*. RAND Corporation.  
[https://www.rand.org/pubs/research\\_reports/RR1993.html](https://www.rand.org/pubs/research_reports/RR1993.html)
- Choi, Y., Mericle, A., & Harachi, T. W. (2006). Using Rasch analysis to test the cross-cultural item equivalence of the Harvard Trauma Questionnaire and the Hopkins

- Symptom Checklist across Vietnamese and Cambodian immigrant mothers. *Journal of Applied Measurement*, 7(1), 16.
- COE - Enrollment Rates of Young Children. (n.d.). Retrieved March 3, 2022, from <https://nces.ed.gov/programs/coe/indicator/cfa>
- Cohen, J. (2006). Social, Emotional, Ethical, and Academic Education: Creating a Climate for Learning, Participation in Democracy, and Well-Being. *Harvard Educational Review*, 76(2), 201-237,285.
- Conrad, K. J., Dennis, M. L., Bezruczko, N., Funk, R. R., & Riley, B. B. (2007). Substance use disorder symptoms: Evidence of differential item functioning by age. *Journal of Applied Measurement*, 8(4), 373.
- Cook, L. L., & Eignor, D. R. (1991). IRT Equating Methods. *Educational Measurement: Issues and Practice*, 10(3), 37–45. <https://doi.org/10.1111/j.1745-3992.1991.tb00207.x>
- Coombs, C. H., Dawes, R. M., & Tversky, A. (1970). *Mathematical psychology: An elementary introduction*.
- Cordier, R., Speyer, R., Chen, Y.-W., Wilkes-Gillan, S., Brown, T., Bourke-Taylor, H., Doma, K., & Leicht, A. (2015). Evaluating the psychometric quality of social skills measures: A systematic review. *PloS One*, 10(7), e0132299.
- Cummings, C., & Haggerty, K. P. (1997). Raising Healthy Children. *Educational Leadership*, 54(8), 28–30.
- Cunningham, J. D., & Bradley, K. D. (n.d.). *Rasch Model to Measure Change*. 33.

- Demaray, M. K., Ruffalo, S. L., Carlson, J., Busse, R., Olson, A. E., McManus, S. M., & Leventhal, A. (1995). Social skills assessment: A comparative evaluation of six published rating scales. *School Psychology Review, 24*(4), 648–671.
- Denham, S. A. (2006). Social-Emotional Competence as Support for School Readiness: What Is It and How Do We Assess It? *Early Education and Development, 17*(1), 57–89. [https://doi.org/10.1207/s15566935eed1701\\_4](https://doi.org/10.1207/s15566935eed1701_4)
- Denham, S. A., Bassett, H. H., Zinsser, K., & Wyatt, T. M. (2014). How preschoolers' social–emotional learning predicts their early school success: Developing theory-promoting, competency-based assessments. *Infant and Child Development, 23*(4), 426–454.
- Dorans, N. J. (1990). Equating Methods and Sampling Designs. *Applied Measurement in Education, 3*(1), 3–17. [https://doi.org/10.1207/s15324818ame0301\\_2](https://doi.org/10.1207/s15324818ame0301_2)
- Dorans, N. J., & Holland, P. W. (2000). Population invariance and the equatability of tests: Basic theory and the linear case. *Journal of Educational Measurement, 37*(4), 281–306.
- Edutopia. (2011). *Social and Emotional Learning: A Short History*. Edutopia. <https://www.edutopia.org/social-emotional-learning-history>
- Evergreen, S. D. H., & Coryn, C. L. S. (2012). Social-Emotional Learning Scale. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 3124–3126). Springer US. [https://doi.org/10.1007/978-1-4419-1428-6\\_409](https://doi.org/10.1007/978-1-4419-1428-6_409)
- Fantuzzo, J., Manz, P. H., & McDermott, P. (1998). Preschool version of the social skills rating system: An empirical analysis of its use with low-income children. *Journal of School Psychology, 36*(2), 199–214.



- Gresham, F., & Elliott, S. (1990). *Gresham: Social skills rating system: Manual*—Google Scholar. [https://scholar-google-com.ezproxy.uky.edu/scholar\\_lookup?title=The%20Social%20Skills%20Rating%20System&publication\\_year=1990&author=Gresham%2CFM&author=Elliott%2CSN#d=gs\\_cit&u=%2Fscholar%3Fq%3Dinfo%3AHQSFbFJmbsIJ%3Ascholar.google.com%2F%26output%3Dcite%26scirp%3D0%26hl%3Den](https://scholar-google-com.ezproxy.uky.edu/scholar_lookup?title=The%20Social%20Skills%20Rating%20System&publication_year=1990&author=Gresham%2CFM&author=Elliott%2CSN#d=gs_cit&u=%2Fscholar%3Fq%3Dinfo%3AHQSFbFJmbsIJ%3Ascholar.google.com%2F%26output%3Dcite%26scirp%3D0%26hl%3Den)
- Gustafsson, J. (1980). Testing and obtaining fit of data to the Rasch model. *British Journal of Mathematical and Statistical Psychology*, 33(2), 205–233.
- Hambleton, R. K. (1989). Principles and selected applications of item response theory. In *Educational measurement, 3rd ed* (pp. 147–200). American Council on Education.
- Hanson, B. A., & Béguin, A. A. (2002). Obtaining a Common Scale for Item Response Theory Item Parameters Using Separate Versus Concurrent Estimation in the Common-Item Equating Design. *Applied Psychological Measurement*, 26(1), 3–24. <https://doi.org/10.1177/0146621602026001001>
- Hawkins, J. D., Catalano, R. F., & Miller, J. Y. (1992). Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: Implications for substance abuse prevention. *Psychological Bulletin*, 112(1), 64.
- Hemmeter, M. L., Ostrosky, M., & Fox, L. (2006). Social and Emotional Foundations for Early Learning: A Conceptual Model for Intervention. *School Psychology Review*, 35(4), 583–601. <https://doi.org/10.1080/02796015.2006.12087963>

- Holland, P. W. (2007). A Framework and History for Score Linking. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and Aligning Scores and Scales* (pp. 5–30). Springer. [https://doi.org/10.1007/978-0-387-49771-6\\_2](https://doi.org/10.1007/978-0-387-49771-6_2)
- Huaqing Qi, C., & Kaiser, A. P. (2003). Behavior problems of preschool children from low-income families: Review of the literature. *Topics in Early Childhood Special Education, 23*(4), 188–216.
- Huba, M. E., & Freed, J. E. (2000). *Learner-centered assessment on college campuses: Shifting the focus from teaching to learning*. ERIC.
- Item Response Theory | Columbia Public Health*. (n.d.). Retrieved September 15, 2021, from <https://www.publichealth.columbia.edu/research/population-health-methods/item-response-theory>
- Jennings, P. A., & Greenberg, M. T. (2009). The Prosocial Classroom: Teacher Social and Emotional Competence in Relation to Student and Classroom Outcomes. *Review of Educational Research, 79*(1), 491–525. <https://doi.org/10.3102/0034654308325693>
- Jones, D. E., Greenberg, M., & Crowley, M. (2015). Early Social-Emotional Functioning and Public Health: The Relationship Between Kindergarten Social Competence and Future Wellness. *American Journal of Public Health, 105*(11), 2283–2290. <https://doi.org/10.2105/AJPH.2015.302630>
- Jones Harden, B., Winslow, M. B., Kendziora, K. T., Shahinfar, A., Rubin, K. H., Fox, N. A., Crowley, M. J., & Zahn-Waxler, C. (2000). Externalizing problems in Head Start children: An ecological exploration. *Early Education and Development, 11*(3), 357–385.

- Jones, S. M., & Bouffard, S. M. (2012). Social and Emotional Learning in Schools: From Programs to Strategies and commentaries. *Social Policy Report*, 26(4), 1–33.  
<https://doi.org/10.1002/j.2379-3988.2012.tb00073.x>
- Katsiyannis, A., Yell, M. L., & Bradley, R. (2001). Reflections on the 25th anniversary of the Individuals with Disabilities Education Act. *Remedial and Special Education*, 22(6), 324–334.
- Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B., & Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success*.
- Kazdin, A. E., & Kendall, P. C. (1998). Current progress and future plans for developing effective treatments: Comments and perspectives. *Journal of Clinical Child Psychology*, 27(2), 217–226.
- Klibanoff, R. S., Levine, S. C., Huttenlocher, J., Vasilyeva, M., & Hedges, L. V. (2006). Preschool children's mathematical knowledge: The effect of teacher" math talk." *Developmental Psychology*, 42(1), 59.
- Kolen, M. J. (1981). Comparison of Traditional and Item Response Theory Methods for Equating Tests. *Journal of Educational Measurement*, 18(1), 1–11.
- Kolen, M. J., & Brennan, R. L. (2014). *Test Equating, Scaling, and Linking: Methods and Practices* (3rd ed.). Springer-Verlag. <https://doi.org/10.1007/978-1-4939-0317-7>
- Kook, S. H., & Varni, J. W. (2008). Validation of the Korean version of the pediatric quality of life inventory™ 4.0 (PedsQL™) generic core scales in school children and adolescents using the Rasch model. *Health and Quality of Life Outcomes*, 6(1), 1–15.

- Landy, S. (2009). *Pathways to competence: Encouraging healthy social and emotional development in young children*. Paul H Brookes Publishing.
- Lehr, F., & Osborn, J. (2005). Helping Your Preschool Child, with Activities for Children from Infancy through Age 5. Revised. In *US Department of Education*. US Department of Education. <https://eric.ed.gov/?id=ED498949>
- Linacre, J. (2003). Dimensionality: Contrasts and variances help for Winsteps Rasch Measurement software. *Daripada* <Http://Www.Winsteps.Com/Winman/Principalcomponents.Htm>.
- Linacre, J. M. (1999). Understanding Rasch measurement: Estimation methods for Rasch measures. *Journal of Outcome Measurement*, 3, 382–405.
- Linacre, J. M. (2000). Comparing “partial credit” and “rating scale” models. *Rasch Measurement Transactions*, 14(3), 768.
- Linacre, J. M. (2012a). Winsteps® Rasch measurement computer program user’s guide. *Beaverton, Oregon: Winsteps. Com*.
- Linacre, J. M. (2012b). Winsteps Rasch Tutorial 2. Retrieved from *Winsteps & Facets Rasch Software Website: <Https://Www.Winsteps.Com/a/Winsteps-Tutorial-2.Pdf>*.
- Linacre, J. M., & Wright, B. D. (2000). Winsteps. URL: <Http://Www.Winsteps.Com/Index.Htm> [Accessed 2013-06-27][WebCite Cache].
- Lombardi, J. (2008). *Time to care: Redesigning child care to promote education, support families, and build communities*. Temple University Press.
- Lord, F. M. (2012). *Applications of item response theory to practical testing problems*. Routledge.

- Magnuson, K. A., Meyers, M. K., Ruhm, C. J., & Waldfogel, J. (2004). Inequality in Preschool Education and School Readiness. *American Educational Research Journal*, 41(1), 115–157. <https://doi.org/10.3102/00028312041001115>
- Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, 47(2), 149–174. <https://doi.org/10.1007/BF02296272>
- McCabe, P. C., & Altamura, M. (2011). Empirically valid strategies to improve social and emotional competence of preschool children. *Psychology in the Schools*, 48(5), 513–540.
- McClelland, M. M., Tominey, S. L., Schmitt, S. A., & Duncan, R. (2017). SEL Interventions in Early Childhood. *The Future of Children*, 27(1), 33–47.
- Melnick, H., Cook-Harvey, C. M., & Darling-Hammond, L. (2017). Encouraging Social and Emotional Learning in the Context of New Accountability. *Learning Policy Institute*.
- Merrell, K. W. (1994). *Preschool and Kindergarten Behavior Scales. Test Manual*. Clinical Psychology Publishing Company, Inc. <https://eric.ed.gov/?id=ED378210>
- Merrell, K. W. (2011). *Social emotional assets and resilience scales*. PAR Lutz, FL.
- Michell, J. (2014). *An introduction to the logic of psychological measurement*. Psychology Press.
- Mueller, C. E., & Bradley, K. D. (2009). Utilizing the rasch model to develop and evaluate items for the tacit knowledge inventory for superintendents (TKIS). *International Journal of Knowledge Management (IJKM)*, 5(3), 73–93.

- Murano, D., Sawyer, J. E., & Lipnevich, A. A. (2020). *A Meta-Analytic Review of Preschool Social and Emotional Learning Interventions—UK Libraries-University of Kentucky (UKY)*. <https://saalck-uky.primo.exlibrisgroup.com>
- Nardi, P. M. (2018). *Doing Survey Research: A Guide to Quantitative Methods* (4th ed.). Routledge. <https://doi.org/10.4324/9781315172231>
- National Research Council. (2000). *From neurons to neighborhoods: The science of early childhood development*.
- Network, N. E. C. C. R. (2005). Early child care and children's development in the primary grades: Follow-up results from the NICHD Study of Early Child Care. *American Educational Research Journal*, 42(3), 537–570.
- Ozer, E. J. (2007). The Effects of School Gardens on Students and Schools: Conceptualization and Considerations for Maximizing Healthy Development. *Health Education & Behavior*, 34(6), 846–863. <https://doi.org/10.1177/1090198106289002>
- Peabody, M. R. (2020). Some Methods and Evaluation for Linking and Equating with Small Samples. *Applied Measurement in Education*, 33(1), 3–9. <https://doi.org/10.1080/08957347.2019.1674304>
- Peabody, M. R., O'Neill, T. R., & Peterson, L. E. (2017). Examining the Functioning and Reliability of the Family Medicine Milestones. *Journal of Graduate Medical Education*, 9(1), 46–53. <https://doi.org/10.4300/JGME-D-16-00172.1>
- Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., & Yazejian, N. (2001). The Relation of Preschool Child-Care Quality to Children's Cognitive and Social Developmental Trajectories through

Second Grade. *Child Development*, 72(5), 1534–1553.

<https://doi.org/10.1111/1467-8624.00364>

Petersen, N. S., Cook, L. L., & Stocking, M. L. (1983). *Irt Versus Conventional Equating Methods: A Comparative Study of Scale Stability*—Nancy S. Petersen, Linda L. Cook, Martha L. Stocking, 1983. <https://journals-sagepub-com.ezproxy.uky.edu/doi/abs/10.3102/10769986008002137>

*Preschool Education* | *Encyclopedia.com*. (n.d.). Retrieved March 14, 2022, from <https://www.encyclopedia.com/social-sciences-and-law/education/education-terms-and-concepts/preschool-education>

Raikes, H., Alexander Pan, B., Luze, G., Tamis-LeMonda, C. S., Brooks-Gunn, J., Constantine, J., Banks Tarullo, L., Abigail Raikes, H., & Rodriguez, E. T. (2006). Mother–child bookreading in low-income families: Correlates and outcomes during the first three years of life. *Child Development*, 77(4), 924–953.

Rasch, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*.

Reardon, S. F. (2018). *The widening academic achievement gap between the rich and the poor*. Routledge.

Rose-Krasnor, L. (1997). The nature of social competence: A theoretical review. *Social Development*, 6(1), 111–135.

Royse, D., Thyer, B. A., & Padgett, D. K. (2015). *Program evaluation: An introduction to an evidence-based approach*. Cengage Learning.

- Schmitt, N., & Stuitts, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement*, 9(4), 367–373.
- Schultz, B. L., Richardson, R. C., Barber, C. R., & Wilcox, D. (2011). A Preschool Pilot Study of Connecting with Others: Lessons for Teaching Social and Emotional Competence. *Early Childhood Education Journal*, 39(2), 143–148.  
<https://doi.org/10.1007/s10643-011-0450-4>
- Spieker, S. J., Larson, N. C., Lewis, S. M., Keller, T. E., & Gilchrist, L. (1999). Developmental trajectories of disruptive behavior problems in preschool children of adolescent mothers. *Child Development*, 70(2), 443–458.
- Stormont, M. (1998). Family factors associated with externalizing disorders in preschoolers. *Journal of Early Intervention*, 21(3), 232–251.
- Stormont, M. (2002). Externalizing behavior problems in young children: Contributing factors and early intervention. *Psychology in the Schools*, 39(2), 127–138.
- Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2004). *The effective provision of pre-school education (EPPE) project technical paper 12: The final report-effective pre-school education*.
- Tietze, W., Hundertmark-Mayser, J., Rossbach, H., Kurmm, V., Wetzels, G., Palacios, J., & Lera, M. (1999). School-age assessment of child development: Longterm impact of pre-school experiences on school success and family-school relationships. URL: [Http://Www. Uni-Bamberg.de/Fileadmin/Uni/Fakultaeten/Ppp\\_lehrstuehle/Elementarpaedagogik/Team/Rossbach/Ecce\\_Study\\_Group.Pdf](http://www.uni-bamberg.de/Fileadmin/Uni/Fakultaeten/Ppp_lehrstuehle/Elementarpaedagogik/Team/Rossbach/Ecce_Study_Group.Pdf) [16.06. 2014].



- Torres, M. M., Domitrovich, C. E., & Bierman, K. L. (2015). Preschool interpersonal relationships predict kindergarten achievement: Mediated by gains in emotion knowledge. *Journal of Applied Developmental Psychology, 39*, 44–52.
- Warren, H., Bassett, H., & Denham, S. (2008). Early emotional development, social cognition and relationships with others. *Zero to Three, 32–39*.
- Wigelsworth, M., Humphrey, N., Kalambouka, A., & Lendrum, A. (2010). A review of key issues in the measurement of children's social and emotional skills. *Educational Psychology in Practice, 26*(2), 173–186.
- Wingersky, M. S., Cook, L. L., & Eignor, D. R. (1987). Specifying the Characteristics of Linking Items Used for Item Response Theory Item Calibration<sup>1,2</sup>. *ETS Research Report Series, 1987*(1), i–100. <https://doi.org/10.1002/j.2330-8516.1987.tb00228.x>
- Wingersky, M. S., & Lord, F. M. (1984). *An Investigation of Methods for Reducing Sampling Error in Certain IRT Procedures—Marilyn S. Wingersky, Frederic M. Lord, 1984*. <https://journals-sagepub-com.ezproxy.uky.edu/doi/abs/10.1177/014662168400800312>
- Wittmer, D., Doll, B., & Strain, P. (1996). Social and emotional development in early childhood: The identification of competence and disabilities. *Journal of Early Intervention, 20*(4), 299–318.
- Wright, B. (1988). The efficacy of unconditional maximum likelihood bias correction: Comment on Jansen, van den Wollenberg, and Wierda. *Applied Psychological Measurement, 12*(3), 315–318.
- Wright, B. (1994). Reasonable mean-square fit values. *Rasch Meas Trans, 8*, 370.

- Wright, B. D. (2003). *Rack and Stack: Time 1 vs. Time 2: Repeated Measures*.  
<https://www.rasch.org/rmt/rmt171a.htm>
- Wright, B. D., & Linacre, J. M. (1989). Observations are always ordinal; measurements, however, must be interval. *Archives of Physical Medicine and Rehabilitation*, 70(12), 857–860.
- Wright, B., & Masters, G. (1982). Rating scale analysis. *Measurement and Statistics*.  
<https://research.acer.edu.au/measurement/2>
- Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., Ludwig, J., Magnuson, K. A., Phillips, D., & Zaslow, M. J. (2013). Investing in Our Future: The Evidence Base on Preschool Education. *Society for Research in Child Development*.
- Yu, C., & Osborn-Popp, S. (2019). Test Equating by Common Items and Common Subjects: Concepts and Applications. *Practical Assessment, Research, and Evaluation*, 10(1). <https://doi.org/10.7275/68dy-z131>

## VITA

### **Education**

#### **University of Kentucky**

Ph.D. Student, Education Policy Studies & Evaluation  
Cahir: Dr. Kelly D. Bradley Dec 2022

M.S, Financing May 2021  
M.S Accounting July 2017

#### **University of Louisville**

B.S Accounting Dec 2016

#### **Harbin Normal University**

B.A Media Production July 2012

### **Research Experience**

#### **University of Kentucky Evaluation Center**

Research Assistant Aug 2018- Dec 2022

### **Publications**

**Li, N., Sampson, S.** “A case study: Rasch stacking comparison of Teacher Efficacy”  
MWER (Mid-Western Educational Research Association), Poster presentation 2021

Niu, C., Bradley, K., Sampson, S., Jin, R., Xia, Y., Shen, L., Zhang, J., Wu, R. & **Li, N.**  
“Simulation study: Evaluating Rater Category Ordering with the JML-Rasch-MFRM  
Model in Facets”  
NCME (National Council on Measurement in Education), poster presentation 2021

Niu, C., **Li, N.**, Toland, M. D. “A Simulation Study: Comparing FDR Correction Methods  
in Using Rasch Trees Modeling for DIF (DIFFERENTIAL ITEM FUNCTIONING)  
Detection”  
IOMW, Poster presentation 2020

Lingat, J. M., Mantooth, R., Dueber, D. M., **Li, N.**, & Toland, M. D. “Detecting gender-based & across campus differential item functioning on a sense of belonging scale in higher education”

APA (American Psychological Association), Poster presentation 2020

Hayden, J. L., **Li, N.**, Parsons, J. & Sampson, S., (2021). (rep.). The Central Appalachian Regional Education and Research Center (CARERC) Current Student Focus Group (pp. 1–14). Lexington, KY. 2022