1-1-2019

Rasch Analysis of Relational Well-Being Within the National Survey of Adoptive Parents of 2007: A Comparison of Multidimensional, Consecutive, and Unidimensional Approaches to Measure Construction

Michael A. Furno
University of Denver

Rasch Analysis of Relational Well-Being Within the National Survey of Adoptive Parents of 2007: A Comparison of Multidimensional, Consecutive, and Unidimensional Approaches to Measure Construction

Abstract

A unidimensional Rasch approach was used to explore whether the data collected through the National Survey of Adoptive Parents of 2007 (NSAP) for the well-being items represented a single latent construct and to establish a base model for comparison. A consecutive approach was then used as an exploratory tool to draw out potential multiple dimensions. Finally, multidimensional item response theory (MIRT) was used to confirm the results of the consecutive approach findings while comparing with the unidimensional baseline. Items within the survey were evaluated for scale function as well as invariance.

The comparison of three approaches (unidimensional, combined consecutive, and 2-dimensional MIRT) found that the combination of Consecutive Dimensions A and B yielded the best fitting model for these data sets. The nested 2-dimensional MIRT model showed better fit than the unidimensional model, but concerns with item position and inconsistent error terms supported the combined consecutive model.

The use of IRT and MIRT analysis techniques helped strengthen the survey by identifying items within the survey that relate to identified constructs. The comparison of three approaches provides practitioners with an example of how to use a consecutive approach in Rasch for exploratory purposes when dimensionality has not already been established.

The NSAP survey was developed to gather data from a large cross-section of adoptive parents in the United States. The well-being subsection gathered data on the parent-child relationship with the intent to assist adoption practitioners, policy-makers, and researchers. Since only twelve of the thirty-nine items were utilized within the models, the data collection opportunity was not fully captured. This lost opportunity of data collection supported the idea of survey development partnerships between topic content experts and psychometricians, when building measures, to maximize the effectiveness of the tool as well as the data gathered.

Document Type

Dissertation

Degree Name

Ph.D.

Department

Quantitative Research Methods

First Advisor

Kathy Green, Ph.D.

Second Advisor

Antonio Olmos, Ph.D.

Third Advisor

Maria Riva, Ph.D.

Keywords
Adoption, Item response theory, Psychometrics, Rasch

Subject Categories
Life Sciences | Research Methods in Life Sciences

Rasch Analysis of Relational Well-Being within the National Survey of Adoptive Parents of 2007: A Comparison of Multidimensional, Consecutive, and Unidimensional Approaches to Measure Construction

A Dissertation

Presented to

the Faculty of the Morgridge College of Education

University of Denver

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

by

Michael A. Furno

June 2019

Advisor: Dr. Kathy E. Green

Author: Michael A. Furno
Title: Rasch Analysis of Relational Well-Being within the National Survey of Adoptive Parents of 2007: A Comparison of Multidimensional, Consecutive, and Unidimensional Approaches to Measure Construction
Advisor: Dr. Kathy E. Green
Degree Date: June 2019

Abstract

A unidimensional Rasch approach was used to explore whether the data collected through the National Survey of Adoptive Parents of 2007 (NSAP) for the well-being items represented a single latent construct and to establish a base model for comparison. A consecutive approach was then used as an exploratory tool to draw out potential multiple dimensions. Finally, multidimensional item response theory (MIRT) was used to confirm the results of the consecutive approach findings while comparing with the unidimensional baseline. Items within the survey were evaluated for scale function as well as invariance.

The comparison of three approaches (unidimensional, combined consecutive, and 2-dimensional MIRT) found that the combination of Consecutive Dimensions A and B yielded the best fitting model for these data sets. The nested 2-dimensional MIRT model showed better fit than the unidimensional model, but concerns with item position and inconsistent error terms supported the combined consecutive model.

The use of IRT and MIRT analysis techniques helped strengthen the survey by identifying items within the survey that relate to identified constructs. The comparison of three approaches provides practitioners with an example of how to use a consecutive

approach in Rasch for exploratory purposes when dimensionality has not already been established.

The NSAP survey was developed to gather data from a large cross-section of adoptive parents in the United States. The well-being subsection gathered data on the parent-child relationship with the intent to assist adoption practitioners, policy-makers, and researchers. Since only twelve of the thirty-nine items were utilized within the models, the data collection opportunity was not fully captured. This lost opportunity of data collection supported the idea of survey development partnerships between topic content experts and psychometricians, when building measures, to maximize the effectiveness of the tool as well as the data gathered.

Acknowledgement

I would like to thank the Research Methods and Statistics faculty and Morgridge College of Education for the opportunity to be a member of the RMS program, with particular thanks to Dr. Ben Siebrase as a fellow classmate, colleague, and friend.

I would like to extend a sincere thank you to my dissertation advisor, Dr. Kathy Green, for all of her encouragement and guidance. Without Dr. Green's constant support, I would not have finished. I want to thank Dr. Antonio Olmos, Dr. Maria Riva, and Dr. Corinne Lengsfeld for their time and insight throughout this project.

I thank my family and friends for their continued support during the pursuit of my degree, notably Eric Reeves and Dr. Casey Dinger. Eric has backed my efforts in this program from the beginning, while keeping me accountable over the years. Casey has been a sounding board and sojourner with me along the way.

Thank you to my children, Matt, Abi, and Dal for their support and sacrifice. Thank you to my parents for the prayers and encouragement. A special thank you to my wife, Tori, for enduring this long journey with patience and positivity.

Table of Contents

List of Tables

List of Figures

Chapter One: Introduction and Review of Literature

Three item response theory approaches using the Rasch model (Rasch, 1960/1980) were compared in the present study: a unidimensional approach, a consecutive approach, and a multidimensional approach to identify dimensionality, invariance by adoption type, and the overall best fitting model for the National Survey of Adoptive Parents (NSAP) well-being item subset. A Rasch model analysis was used to identify whether multiple dimensions were present in the data and, then, which items fit with each dimension. Knowing more precisely the dimensionality of a measure can strengthen researchers' theories by effectively identifying the best fitting items/dimensions for their research questions.

The three Rasch approaches provide value though their prescribed process to scale development, model fit, and interpretation. The unidimensional is the traditional approach for Rasch analysis and identifies how well items fit a single theme. The unidimensional approach assumes all survey items are connected to the same single construct when determining model fit. Adjustments to the items and response scale are made to improve the model fit to better measure the construct. The single focus of the unidimensional approach allows for easier interpretation of the results. The consecutive approach begins with same assumption of the unidimensional approach, a single latent construct, while exploring the possibility of more dimensions or latent constructs that separate the items into smaller separate groups. Utilizing the consecutive approach equips the researcher with a process to examine unexplained error and items not reflecting the first construct. From misfitting items, potential new dimensions emerge. As more dimensions emerge through the consecutive approach, more overall variance is explained and precision is added to the measures. The consecutive approach produces, if additional dimensions are found, an emergent number of unidimensional item clusters with model fits determined for each individual cluster and does not have an overarching model fit incorporating all dimensions. Often the consecutive approach is used to examine the parameter estimates of individual sub-clusters within existing multidimensional models. Drawbacks of the consecutive approach include an inability to establish associations between the dimensions within the measurement framework (though certainly dimension correlations can be calculated when dimensions are established) and overestimation of measurement error on the items and persons for the examined dimension (Baghaei & Grotjahn, 2014; Huang, Wang, Chen, & Su, 2013).

The multidimensional approach is the most complicated of the three approaches, since this approach examines the known or perceived dimensions within the survey as well as relational connections between the dimensions when determining model fit. The multidimensional approach provides a more complicated yet comprehensive perspective of a multidimensional model. While the dimension estimates within the consecutive approach are more easily interpretable, these findings are narrower in scope (Wiberg, 2012). Like the unidimensional approach, the multidimensional approach focuses less on the emergent dimensions, as with the consecutive approach, and more on overall model fit.

Comparing these three approaches utilizing a single dataset allows the researcher to appreciate the insight each process provides as well as determining the best fitting model for the data at hand. While all three approaches are commonly used in practice, there are few reports comparing the three approaches that can be used to guide practitioners in analysis of the structure of their measures. These three approaches were employed with the well-being subset of items from the National Survey of Adoptive Parents (NSAP), a survey which has seen relatively little psychometric work.

At the time of the NSAP administration in 2007, approximately 1.8 million adopted children lived in the United States (Child Health U.S.A., 2010). Of these children, 25% had been adopted internationally, 37% through the foster care system, and 38% through private domestic arrangements (Harwood, Feng, & Yu, 2013; Vandivere, Malm, & Radel, 2009). Although there are perceived similarities between adoptions, each adoption is unique because there is no single path to becoming an adoptive family. Each member of the adoptive family contributes experiences and characteristics that affect the family's existence. Characteristics such as the age at which the child was adopted, the physical health of the child, the fertility of the parent, the parent's rationale for adoption, experiences including the nurturing a child received, abuse the child may have experienced, and/or the parent's ability to build close relationships influence child and family development. The effect of the collected experiences and characteristics differ for each person based upon intensity and duration of the situation. All aspects of the individuals involved contribute to the mosaic of the adoptive family. However, the uniqueness of each adoption story has hindered researchers' ability to control the characteristics and experiences brought by the participants to their studies, which limits the generalizability of the research findings and so limits direction for support and intervention.

The success of the parent-child relationship determines the overall success of an adoption (Good, 2015; Neil, 2012; Zamostny, O'Brien, Baden, & Wiley, 2003). As a part of the 2000 U.S. Census through the National Survey of Children's Health (NSCH) in 2007, it was determined that data on adoptive families should be gathered to establish national estimates of adoptive children and their families' well-being, health, and other characteristics. The NSAP utilized the identified adoptive children within the NSCH to provide a random representative sample in order to collect the desired national estimates on the well-being of adoptive children and their families.

The NSAP contains subsections that examine topics including demographics of adoptive families, financing foster-to-adopt, parent-child relational well-being, etc. Researchers have used the well-being section of the NSAP to show the importance of the parent-child relationship and the strength of relationships of the participating families; however, the psychometric quality of this subsection has not been established. Use of an instrument with limited supporting psychometric information has limited the effectiveness of research, as reliability, validity, and generalizability of measures as solid data sources are unknown. An important step in evaluating an instrument is to identify the dimensional structure of the instrument. Understanding the dimensionality allows the researcher to identify more clearly the construct assessed by each dimension. The dimensionality of the NSAP well-being subset is unclear.

Two studies examining the dimensions of the relational well-being subset of the NSAP have been conducted. The first used exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), while the second study used principal components analysis (PCA) and CFA. The first study revealed both a one-factor model and a two-factor model fitting the data with the two-factor model fitting best. Similarly, the second study produced a two-factor model, which utilized comparable items. These studies support a multidimensional subsection structure; however, these studies revealed substantial missing data generated by the data collection process. The first study used pairwise deletion for the missing data and the second study imputed the missing data using a multiple expectation-maximization algorithm. Both of the studies limited the items to eight and six respectively, which is less than a quarter of the total questions within the well-being subset (Park, Barth, & Harrington, 2013). The trimming process from forty-nine to six and then to eight items limits the ability to draw conclusions about the whole subsection when so few items were included. Using a Rasch analysis approach addresses some of the concerns found in analyses such as CFA, EFA, and PCA, and can help to confirm the dimensionality of the relational well-being items through the use of a different measurement model. Classical test theory treats the measure as a whole, while the Rasch analysis examines the contribution and fit of items individually. An item-focused analysis estimates the standard error per item, which provides insight into the amount of item contribution to the measure or dimension. Further, Rasch analysis relies on the contribution of individual items to the overall measure irrespective of missing observations; thus, missing data are accommodated readily in a Rasch analysis. This analysis strengthens researchers' understanding of the data and use of these data by helping to select optimal items for their intended analysis.

The results of this research study assist practitioners to focus on items from the NSAP that best represent relational well-being, and more effectively provide services to strengthen the parent-child relationship that ultimately lead to stable child placement decisions and permanency in the placement. In addition, the results of these findings reinforce future research by estimating measurement reliability and validity. The examination of the three approaches provides practitioners with examples of how to group items from within the well-being subset of this survey to yield more psychometrically sound results.

National Survey of Adoptive Parents (2007)

With just under 2% of the United States' children being adopted, it was decided by the United States Department of Health and Human Services through the Centers for Disease Control that the adoptive community and United States government needed to gain better insights into the general characteristic and resource needs of the population. To fulfill this need, the NSAP (National Survey of Adoptive Parents) was created and conducted (Vandivere & McKlindon, 2010). A survey on a nationwide scale had not previously been conducted with a focus on the entire adoptive community in the United States. In 2005, the Urban Institute and National Opinion Research Center (NORC) of the University of Chicago were issued a task ordered by the Assistant Secretary for Planning and Evaluation (ASPE) to develop an instrument for the National Survey of Adoptive Families. The groups conducted a literature review of adoptive research and past adoption surveys. The findings and existing items were categorized into applicable topics and each topic was given a level of importance. From this pool of items, an initial survey was formulated and reviewed by ASPE. Suggestions ranging from new topics, wording of questions, and approaches to inter-country or international adoption were reviewed and changes were made to enhance the survey. Next, seven adoptive parents (five foster-to-adopt parents, one private domestic parent, and one inter-country parent) provided feedback through cognitive interviews to determine how well they understood each of the survey questions. In the last step, using the final draft of the survey, eight adoptive parents (two foster-to-adopt parents, three private domestic parents, and three inter-country parents) participated in a pretest to assess the survey for flow and time needed for completion (Bramlett et al., 2010).

The NSAP administration established a large national sample of data with 2,089 participants, including various types of adoptions. This survey was more representative than other adoption surveys. Past adoption research used small sample sizes focused on particular populations with few studies examining all types of adoption. A more recent adoption survey, the National Adoptive Families Study (2012), with 437 participants, focused on foster-to-adopt families and dissolution, but was not representative of all adoption types (Hartinger-Saunders, Trouteaud, & Matos-Johnson, 2014). Additionally, the NSAP focused a section of the survey on addressing the well-being of the parent-child relationship (Harwood et al., 2013; Vandivere & McKlindon, 2010). Quality of the relationship was measured through the NSAP by examining the closeness of the parent and child, reported child affection, and the parent's satisfaction with the relationship, which helped to acknowledge the need for adoptive family support and to identify future research needs (Harwood et al., 2013).

Results of analyses of the well-being subsection of the NSAP have been reported in a number of journal articles. Vandivere and McKlindon (2010) divided well-being into three factors as a way of analyzing well-being within the participant population, with two or three items placed into each factor; however, no factor analysis was reported or explanation provided regarding why these factors with these particular items were selected for their analysis. Another article used data from both the NSAP and the 2007 National Survey of Children's Health (the parent survey of the NSAP) to examine differences between types of adoptive families in the areas of demographic characteristics and health and well-being (Radel, Bramlett, & Waters, 2010). This article did not support or refute the well-being category as a construct.

Another study using NSAP data found there was no difference between private adoptions and foster-to-adopt in parent-child relationship quality. Child characteristics such as being a boy, older age at placement, older age, healthcare special needs, or fourth of five children, did contribute to lower parent-child relationship quality scores. In addition, a non-contributing factor to lower parent-child relationship quality scores was the presence of biological children within the family (one of two family characteristics (Socioeconomic Status (SES) is the other)). Little impact on relationship quality was found for pre-adoption adversity (fifth child characteristic). Pre-adoption adversity may have had less of an impact on the parent-child relationship quality as demonstrated in the well-being section of the NSAP. The impact may have been lessened because the adoptive parents sought out assistance, they may have had lowered expectations of the relationship, and/or the NSAP data may be flawed through the parents' limited knowledge of actual experience when responding to the question (Tan, Major, Marn, Na, & Jackson, 2015).

Country of origin was used to evaluate the parent-child relationship through a multiple regression analysis. The study found that country of origin was the sole statistically significant predictor of well-being when comparing United States private adoption and adoption from other countries. It was found that the well-being of the parent-child relationship in private domestic adoptions were stronger than the relationships of international adoptions. There were some fluctuations when comparing the well-being of the parent-child relationship between non-U.S. countries, inter-country adoption, and U.S. foster-to-adopt families. Other predictors included in the model were pre-adoption adversity, age at placement, gender, and special health needs (Tan et al., 2015). This study did not actually use data from the well-being subset but did infer relational health of the parent-child relationship.

The weakness of large-scale surveys on adoption has been that although adoptions are similar in general, the lived experiences of the children are very different. A variety of factors, such as age at placement, type of adoption, the presence and type of trauma, the child's personality, and his/her handling of the loss of birth parents, individually impact the adoptee's interpretation of the survey questions, and this has made it difficult to generalize findings (Miller, Fan, & Grotevant, 2005). The NSAP has used the perspectives of the adoptive parents, which limits a complete understanding of the relationship to only one side of the relationship. Despite the one-sided nature of asking only the parents, it has been commonly understood that parents tend to provide more valid responses than children (Miller et al., 2005). Often with large scale surveys, it has been difficult to ensure that the subject met the survey inclusion criteria because of the complexity in verification (Miller et al., 2005). The NSAP protocol tried to minimize this concern by utilizing U.S. Census data and processes. Large national samples are understood as better for generalization than smaller nonprobability samples, which have been more common in adoption research due to the rarity of adoption as well as the specialized research/practitioner settings (Miller et al., 2005).

Statement of the Problem

The National Survey of Adoptive Parents 2007 collected data from the largest cross-section of adoptive parents in the United States at that time, which has made the findings invaluable to the adoptive community. This national study provided data to adoption researchers and practitioners with limited resources. The NSAP parent-child well-being subsection tried to capture a measure of the degree of relational well-being between the adoptive parent and child; however, it was unclear if the measure was identifying a single dimension of well-being or multiple dimensions of well-being. Better understanding of what the measure identifies will assist future researchers and practitioners in their work when using this measure. Well-being items were designed to explore the parent-child relationship and the influencing factors between the adoptive parent and the adoptee in order to support the child during their developmental years and early stages of the adoptive relationship (Vandivere, Malm, & Radel, 2009). Utilizing the data from the NSAP, the current study applied item response theory to the parent-child relational well-being subsection to provide insight into the structure and psychometric quality of the survey. Without estimated reliability and validity, survey data analyses lack consistency and focus. Appraisal of the structure and estimates reliability and validity aids researchers and adoption practitioners effectively in their research pursuits with the data.

Application of the three item response theory approaches (unidimensional, consecutive, and multidimensional) with the NSAP well-being items allowed for insight into the value of each approach. Since this survey was developed through multiple phases, with a variety of groups contributing to the survey items, and, since the overall size of the survey was extensive, the precision of the subsets within the NSAP was open to examination. The use of item response theory (IRT) models examined the structure through the probability of endorsing scores on the items and a person's probability of agreeing with the item. IRT helped to verify consistency of the responses to the scale used for each item. The NSAP well-being item subset accommodated the comparison of the three item response theory approaches: unidimensional, consecutive, and multidimensional.

The unidimensional approach allowed for a concentrated analysis of model fit to a single theme. This approach was used to determine if the NSAP well-being subset fit a single construct, presumably parent-child relational well-being. The consecutive approach allowed for exploration of dimensionality when the analysis suggested more than one dimension. There were few articles located where the consecutive approach was used and few that provide a comparison of the three approaches. Dimensionality is found through a stepwise process when using the consecutive approach. The consecutive approach has often been used as a way to examine the fit of sub-scales within multidimensional models. The use of the consecutive approach opened the analysis to an increased possibility of estimation error. Using a multidimensional approach has the effect of reducing measurement error. Employing the multidimensional approach allowed an examination of model fit of the known or perceived dimensions and the relationship between the dimensions (Briggs & Wilson, 2003; Wiberg, 2012).

Purpose of the Study

The primary purpose of this study was to compare the use of three item response theory approaches to measure construction: unidimensional, consecutive, and multidimensional, with an exemplar dataset.

The NSAP authors indirectly suggested that subsections contribute to a single overarching construct, which could be interpreted as supporting a unidimensional construct. An initial review of the parent-child well-being subsection items might lead to the hypothesis that this subsection is unidimensional, as all items represent the well-being construct. However, many researchers have used the findings of this survey to draw conclusions about parent-child relational well-being by selecting individual items within the survey as representative of well-being or of other constructs. Few psychometric analyses have been conducted on the well-being subsection of this measure. A key study recommends the use of structural equation modeling to identify dimensionality and potential difference between adoption types (Harwood et al., 2013); however, due partly to extensive incomplete data, early confirmatory factor analyses of the data have reduced the instrument item subset to being nearly unusable as a single construct instrument (Park et al., 2013).

Due to these limitations in prior research, responses to this measure were analyzed using item response theory, comparing the fit of a measure developed using a unidimensional Rasch model to results using a consecutive approach, to results using a multidimensional Rasch model. In the presence of sparse data (due to extensive missing data) and a survey created without specific attention to theory, the purpose was to determine whether paring down items to a core unidimensional construct, retaining as many items as possible via identification of independent subconstructs, or allowing subconstruct correlation via multidimensional IRT would yield the most effective solution.

Measurement invariance was addressed through differential item functioning to determine if the responses to the well-being subset were consistent across three types of adoption (inter-country, domestic private, and foster-to-adopt) or if there was measurement bias between groups (Bahraini, 2008; C.C. Chang et al., 2015). Differences in experiences and outcomes have been found between the key types of adoption (inter-country, domestic private, and foster-to-adopt), and since the NSAP has collected data across these adoption types, further analysis of the item function through the lenses of the adoption types would increase the support for validity of the original findings.

The secondary outcome of this study was the identification of the dimensionality and functioning of items measuring parent-child well-being found in the NSAP. This information can prove useful for researchers who address the adoption experience using NSAP data.

Research Questions

1. Did the data from the parent-child well-being subsection found in National Survey of Adoptive Parents (2007), NSAP, support a unidimensional or multidimensional structure when using a Rasch partial credit model for analysis?
    a. Were the psychometric properties of model fit, item fit, and reliability more suitable for the NSAP data within a unidimensional model or multidimensional model?
    b. How did item and person logit positions differ between the unidimensional and multi-dimensional findings?
    c. Did item and person indicators of position differ between software (Winsteps and ConQuest) when using the unidimensional and consecutive approaches?
    d. Which approach yielded a better model fit for the well-being subsection of the NSAP?
2. Were scale response categories used appropriately for each of the included items?
3. Once the dimensionality had been established and the item categories determined, did respondents for different adoption types (inter-country, domestic-private, and foster-to-adopt) interpret the items differently as observed through differential

item functioning (DIF)? How did the differential item functioning results for adoption type compare between the unidimensional, the consecutive, and the multi-dimensional approach?

4. How did the person logit position and item difficulty compare for each dimension found within models from the unidimensional, consecutive, and multidimensional approaches across models?
    a. Did the person logit positions correlate across dimensions/models and software packages?
    b. Using a cluster of three independent variables, Adoptive Family with or without Biological Children, Child Lived with Birth Family, and Adoptive Parent/s and Child of Differing Races, as predictors in a regression analysis, were the R^2 values comparable between the models and the software?
    c. Using the same cluster of family characteristic variables as independent variable in a canonical correlation analysis, were the canonical R^2 values comparable between the models and the software?
    d. How did the item difficulty compare across the models examined?
5. With the best fitting model selected, were the dimensionality and model fit replicated through the use of a second half of the dataset for cross validation? Were the item fit, DIF, and validation measures comparable across the two halves of the dataset?

General Perspective on Adoption

The goal of the child welfare system is to establish home permanency for the child in order to establish a stable and secure environment in which the child can thrive. The primary concern of permanency is to determine which home provides the child with the best environment for stability, while ensuring that the child is not removed permanently from the birth home, if there is a possibility of future stability (Hollinger, 2000; Skivenes & Tefre, 2012). The birth parents' behaviors are evaluated when childcare officials consider removing a child from the home. If it is determined that the child is experiencing undue harm or risk within the birth family home, then the child will be removed to foster care and potentially become eligible for adoption with or without the birth parents' consent (Gilbert, Parton, & Skivenes, 2011; Skivenes & Tefre, 2012). Placing children is a process overseen by social workers, which entails assessments of the prospective home and parents to fit the child in need of a permanent home (Mountjoy & Vanlandingham, 2015).

The removal of a child from the care of the birth parents is viewed as a drastic step and for this reason the adoptive parents' readiness for bringing the child into their home is tested. Once approved, adoptive parents are accepting the responsibility to parent the adoptee, a child not birthed by the prospective parents, on a permanent basis (Mountjoy & Vanlandingham, 2015). Adoptive parents need to be willing and capable of providing a stable environment for the adoptee, within the context of the birth parents', adoptive parents', or child's past experiences (Colonnesi et al., 2013). In the U.S., permanency is a priority both for children within the child welfare system and for those who are outside the U.S. child welfare system (Berrick, 2008). Placing a child into an adoptive family is determined to be a better option due to the establishment of a stable and secure home versus the perceived continued maltreatment. This consistent safe environment allows a child to establish a personal and family identity through a continuous relationship with the adoptive parent/s (Zill & Bramlett, 2014).

There are three basic types of adoption (inter-country, domestic private, and foster-to-adopt). Inter-country or international adoption occurs when the adoptive parents seek a child from a country outside the United States. This process has a number of additional costs and legal steps that need to be fulfilled in order for a child to be legally adopted. A standardized set of expectations is laid out for countries participating with the Hague Convention on Protection of Children to ensure the safety of the children and to follow specified procedures to prevent child trafficking (Lee, 2003). Each country that participates in the inter-country adoption process has cultural and social conditions leading to the resulting adoptions, such as China's former one child rule, which allowed Chinese families to have only one child, or the Confucian influence over South Korea, which emphasizes bloodlines and so limiting Korean domestic adoptions (Tan et al., 2015).

The domestic private adoption occurs when the birth parent/s relinquishes the parenting responsibility or are unable to parent the child (e.g., deceased birth parents) and the adoptive non-related family takes over the legal responsibility of parenting the child within the United States (Vandivere et al., 2009). Often the child is an infant and the adoption occurs within the particular state in which the child was born (Wolfgram, 2008).

Birth parents relinquish a child due to a lack of resources or other conditions preventing them from parenting (Tan et al., 2015).

In a foster-to-adopt situation the child has been removed from the birth parents' custody due to an inability or unwillingness to provide the necessary care (Zill, 2011). Birth parents have either voluntarily relinquished their parental rights or these rights have been legally terminated (Vandivere et al., 2009). Once the birth parents' rights have been removed, the adoptive parents take the necessary steps to complete the adoption process. The child may be living with foster parents while the adoption process proceeds or, in some circumstances, the adoptive parents can serve as the foster parents. In 2007, according to records from the Adoption and Foster Care Analysis and Reporting System (AFCARS), 54% of children entering foster care experienced neglect. Approximately 1% of the children in the United States were in foster care at some point in 2011 (Zill & Bramlett, 2014); and in 2012 about 18% of the children within the foster care system were adopted by non-related parents (Child Welfare Information Gateway, 2013).

Beyond the primary focus of finding and establishing a stable and secure environment to support the child's growth, social workers, the government, and adoptive parents must consider other factors, such as the openness of the adoption. Open adoption is established when a relationship is maintained with one or more of the birth relatives and the adoptive family. The practice of open adoption began in the 1970s (Hoksbergen & ter Laak, 2005) and has become more prominent since the 1980s in part due to the Adoption Assistance and Child Welfare Act of 1980 (Berry, 1998). The amount of contact and information sharing, identified as the openness of the relationship, varies for each relationship and is highly controlled by the adoptive parents (Berry, 1991). The variation in openness has made it difficult to conclusively determine the success of open adoptions; however, many researchers have suggested that openness is a benefit to the adoptive relationship (Berry, 1991; Berry, Dylla, Barth, & Needell, 1998; Grotevant, Ross, Marchel, & Mcroy, 1999). In contrast, Vandivere and Mcklindon (2010) found a slight negative impact on the socio-emotional health of the adoptee in an open adoption. Prior to the 1970s, adoptions were always closed with no contact with the birth parents, as it was believed, this method protected all of the participants, but current findings suggest that the lack of contact leads to unnecessary secrecy (Goodman, Emery, & Haugaard, 1998). The secrecy created by closed adoptions has been called into question by birth mothers and adoptees, which has led to the increase in open adoptions (Goodman et al., 1998; Wolfgram, 2008).

Another consideration for the adoptive community is the matching of racially different children and parents in adoptive families, identified as transracial adoption, and can occur in both domestic and international adoptions (Lee, 2003). Transracial adoptions are the most visually obvious adoptions due to the physical differences and often draw the most attention, both negative and positive. Much of the social and political controversy surrounding adoption focuses on transracial adoption (Zamostny, O'Brien, Baden, & Wiley, 2003). Early policies around domestic transracial adoption in the U.S. met with strong resistance from the Native American community and the National Association of Black Social Workers. Both groups believed that the domestic promotion of transracial adoptions would lead to cultural genocide for minority populations. Due to these concerns, policies quickly changed by giving preference to same-race placements (Lee, 2003; Simon & Altstein, 2000). Transracial international adoptions make up the vast majority of transracial adoptions in the U.S. Besides working through the social and political concerns of the particular countries involved with the adoption, the main concerns with these adoptions are possibilities of child trafficking, forced labor, and cultural imperialism (Engel, Phillips, & Dellacava, 2007; Lee, 2003; Sass, 2014). The Hague Convention on the Protection of Children has brought participating nations together to establish strict standardized conditions in which international adoption can occur. The Inter-Country Adoption Act and Child Citizenship Act of 2000 is an example of a policy that has resulted in the closure of a number of adopting countries to U.S. adoptive parents. The intent was to disallow adoption of children from countries lacking the infrastructure to ensure children are truly available for adoption (Lee, 2003).

While most adoptions that reach the final phase of the adoption process are finalized and forever, some adoptions are not completed or need to be ended. A disrupted adoption is one that does not go to completion. The child is paired with and may even live with the prospective parent/s, but the adoption process is not completed. The prospective parents generally initiate a disrupted adoption; however, a legal questioning of birth parents' relinquishment can disrupt the adoption, social workers' questioning of adoptive parents' readiness, or governmental issues. The dissolution of an adoption occurs when an adoptive parent relinquishes the responsibility of parenting the child, at which point the child returns to the status of orphan and is removed from the adoptive home. In the United States these children are placed in foster care and the adoption process may begin again (Burke, Schlueter, Vandercoy, & Authier, 2014; Vandivere & McKlindon, 2010).

Parent-child Relationship

The parent-child relationship is the most impactful relationship during the child's development. This relationship establishes the environment from which the child draws identity, cognitive development, socialization techniques, and physical health (Harwood et al., 2013). Key aspects of adoption influencing the relationship between the parent and child are the contexts under which the adoption occurred. Adoption specific variables such as congregate care (e.g., institutionalization or group homes), age at placement, prior maltreatment (e.g., prenatal drug and alcohol abuse, physical and sexual abuse, neglect), child's race or lived ethnicity, transracial family, parental attachment, poverty level, and new income level impact the parent-child relationship (Vandivere & McKlindon, 2010). Adoptive children experience higher need as identified through lower achievement and behavioral problems, which are attributed to trauma and loss (Zill & Bramlett, 2014). Every adoptee and many adoptive parents have experienced loss at some level, whether through the loss of their birth parents or through infertility. It is necessary to understand and address the sense of loss experienced and how this loss impacts the new relationship by all involved (Singer & Krebs, 2008). Grieving through the loss is an ongoing process towards the point of acceptance of the new parent-child relationship (Mountjoy & Vanlandingham, 2015).

The environment or contextual conditions of the adoptive child exert influence on the development of the adopted children in the domains of physical growth, attachment, cognitive development and school achievement, self-esteem, and behavior problems (van Ijzendoorn & Juffer, 2006). Research, also, has shown that the stress of adoption as well as the pre-adoption conditions in which the children have been raised increase the likelihood of emotional and behavior problems (Rosnati & Barni, 2008). However, not all researchers agree that the child's experience and conditions prior to adoption are causal but rather show an association between these characteristics and the well-being of the child and the well-being of the parent-child relationship (Vandivere & McKlindon, 2010). The environment we grow up in impacts how our genetics develop either positively or negatively; nurture does influence nature (Rutter, 2005). Maltreatment leads to low self-esteem (Egeland, Sroufe, & Erickson, 1983; Kim & Cicchetti, 2004; Toth et al., 1997). A sense of self-worth and trust in oneself has been linked to a secure base provided by sensitive parents (Sroufe, Egeland, Carlson, & Collins, 2009; van Ijzendoorn & Juffer, 2006)

In the 1930s, John Bowlby, influenced by psychoanalysis, began theorizing about the ties between a mother and her child and the impact of disrupting this relationship on the mother and child. Independently, at first, Mary Ainsworth developed empirical methods to study Bowlby's theory of attachment. As the theory emerged, the two researcher collaborated setting the path for future attachment research (Goldberg, Muir, & Kerr, 1995). Attachment relationships are characterized as either secure or insecure and then further categorized through types of attachment to include secure, ambivalent, avoidant, and disorganized (Rees, 2008). A meta-analysis found that 47% of adopted children were securely attached in comparison 67% of non-adopted children who were securely attached, showing that adopted children present as less securely attached (van Ijzendoorn, Goldberg, Kroonenberg, & Frenkel, 1992; van Ijzendoorn & Juffer, 2006).

The quality of the attachment between the parent and child is impacted by a number of factors such as the child's characteristics, levels of trauma, responsiveness of the caregiver(s), and the duration of these factors (Harwood et al., 2013; Vandivere & McKlindon, 2010). Other factors, including sexual and physical abuse as well as emotional and general neglect, impact a person's capacity to develop securely attached relationships (Carnes-Holt & Bratton, 2014). In addition, the parents' own attachment styles, through their own responses to the relationship, impact how the child attaches to the parent and all other relationships, despite the intentions of the caregivers (Steele, Hodges, Kaniuk, Hillman, & Henderson, 2003).

In a 1952 report to the World Health Organization, Bowlby reported that the institutionalization of children decreases the child's ability to develop "stable and continuous attachment relationships" (van den Dries, Juffer, van Ijzendoorn, & Bakermans-Kranenburg, 2009, p. 411). It was expected that the children from institutional care would suffer more negative developmental effects that result in less responsiveness and attachment, while showing greater indiscriminate friendliness than children do from either foster care or the non-adopted children (van den Dries et al., 2012). The attachment experiences differed based upon the length each child was in the (Chinese) orphanage. The longer the child was institutionalized, the greater the impact on the child's ability to attach with the adoptive mother (Lancaster & Nelson, 2009). The orphanage experience, a loss of culture, and a loss of birth parents have been shown to be precursors for disrupted attachments, behavioral problems, and mental health concerns (Juffer & van Ijzendoorn, 2005; Van den Dries, Juffer, van Ijzendoorn, & Bakermans-Kranenburg, 2009). Children who had been institutionalized were at an increased risk of exhibiting disorganized attachment patterns compared to children remaining in stable birth homes (Van Londen et al., 2007). Institutionalization includes group home, orphanage, and psychiatric placement living (Vandivere & McKlindon, 2010). Institutionalization can produce developmental delays due to negative experiences and environmental conditions (Tan et al., 2015). The longer a child spends in group care, the greater the physical growth delays become (van Ijzendoorn & Juffer, 2006). Also, the impact of institutionalization is ongoing despite post-adoption experiences (Harwood et al., 2013; Vandivere & McKlindon, 2010). Group care or institutionalization inhibit a child's ability to develop empathy and emotional understanding (van Ijzendoorn & Juffer, 2006; Vorria et al., 2006). "The change of environment from impersonal group care of low quality to normal family life is more drastic than in any other large-scale intervention such as Head Start or Sure Start" (van Ijzendoorn & Juffer, 2006), p. 1229).

The stability of the parent-child relationship is critical for the long-term success of the child's development. The parent-child relationship establishes the base of current and future relationships (Woodhouse, Dykas, & Cassidy, 2009). Children who experience interruptions in the relationship with their primary caregiver may find a negative impact on their social, relational, and emotional development (Pace & Zavattini, 2011). The importance of stability at the start of a child's life has been emphasized as a key to building secure attachments. The U.S. Department of Health and Human Services continues to recognize the belief that a stable home, whether with birth family, adoptive family, or foster care, supports a child's sense of safety, permanency, development, and overall sense of well-being. O'Neil, Risley-Curtiss, Ayon, and Rankin-Williams (2012) focused on the importance of children who have experienced trauma being placed in stable environments with the intention of minimizing further trauma.

Data from the National Survey of Children and Adolescent Well-Being (NSCAW) were used to develop a logistic regression model for the purposes of predicting the level of stability for children placed in foster care based upon the characteristics of the foster caregiver and the foster homes. NSCAW was a national survey which contained a subgroup examining long-term foster care (LTFC) children (n = 436). Of the five waves conducted through NSCAW only the first and third were used in the development of the tested models. The researchers examined the impact of the placement stability on the consistency of the caregiver between the data collection waves, caregiver characteristics, characteristics of the child, as determined by, the Child Behavior Checklist (CBCL) and the Social Skills Rating System (SSRS), and the caregiver-child relationship or emotional support within the relationship. The final logistic regression model using caregiver characteristics was not statistically significant with $p = .09$. This model started with ten characteristics and through stepwise trimming ended with four characteristics: caregiver race, placement type, number of household members, and caregiver's experience. It was found that there was no significant impact on the placement stability of characteristics of the child; however, the child's race effect remained stable through the modeling with $p = .07$. The placement type significantly affected placement stability (O'Neill, Risley-Curtiss, Ayón, & Williams, 2012). This study did not explore the quality of the caregiver-child relationship or the caregiver-birth parent relationship, which were needed next steps.

Further research then found that the quality and struggles of the parent-child relationship influenced the placement stability. Instability within a home (birth home, adoptive home, or foster home) can result in a disruption, which creates the child's feelings of loss, anxiety, and depression; and disruptions can influence the child's socioemotional development, trust relationships, behavior, and academic success. Additionally, research found that children with behavioral problems and mental health issues were more likely to have experienced disruption from foster homes. Disruption occurs when a child is removed from his/her residence. Foster homes with more children experienced more disruptions, especially with the most newly placed child into the home (Tan et al., 2015; Vandivere & McKlindon, 2010).

According to Mountjoy and Van Landingham (2015), "the higher the levels of stability and security within the home expressed through the emotional, social, and relational maturity each adult displays increases the potential for stability and security within the lives of each of their foster or adopted children" (p. 12). It is important for both the adoptive parents and the child to work through their past experiences and sense of loss in order to establish more secure relationships in the future (Singer & Krebs, 2008). Since the parent-child relationship develops throughout life, it is important for the caregiver to model appropriate behavior/parenting (Mountjoy & Vanlandingham, 2015). The stability of the home environment and the child's relationships are greatly impacted by the experiences and knowledge of the parents (Harwood et al., 2013). Stability supports well-being through emotional development and socialization (Zill & Bramlett, 2014). The child is dependent on the parent to create an environment that supports healthy relational development (Woolgar & Scott, 2013), so it is necessary to understand the conditions, experiences and resources that can disrupt the environment. How the parent came to the decision to adopt, as well as the type of resources available to the family, affects the overall context regarding the child placement.

Once a child has been placed within a home, parenting stress can destabilize the home environment. One study found that stress for the adoptive parents increased with male adoptees, as the age at adoption increased, or with children with special needs status (Palacios & Sanchez-Sandoval, 2005). In contrast, a 2010 study of international adoption of children with a mean age below eighteen months found that there was no significant relationship between the age of the child, special needs status of the child, or gender of child to parenting stress. These findings conflict with previous research which could be due to the age differences between the studies (Viana & Welsh, 2010). According to Judge (2004), the ability of the child to attach to a caregiver decreases in direct relationship to the increasing length of institutionalization, which increases the stress within the home. The age of the child at the time of adoption and the length of institutionalization are highly correlated, which makes it difficult to separate the impact of institutionalization and the age of the child at adoption on the home (Judge, 2004). These discrepancies in the research may suggest the need to focus more on the relational interactions than on child or parent characteristics (Viana & Welsh, 2010).

Stress heightens the difficulty of child rearing, the success of the adoption, and parents' satisfaction with the adoption. Their satisfaction is tied to their preparedness for the conditions of the adoption and the adoptive experience meeting their expectations. The less stress the adoptive parents experience, the more satisfaction they have about the adoption. A better adoptive experience for the child is possible through the increased stability of the home, open communication, and a positive view of adoption (Palacios & Sanchez-Sandoval, 2005). Mothers who perceived greater levels of post-adoption support reported higher levels of satisfaction with adoption and lower levels of stress (Viana & Welsh, 2010). Stresses may be due to isolation or depression. Higher stress has been tied to a decrease in attachment (Judge, 2004). Lower stress helps parents encourage children develop the skills to attach (C. D. M. Brodzinsky, Smith, & Brodzinsky, 1998).

An aspect of the stability provided through the family environment and the child's experiences contribute to the degree to which the child and parent form attachments. A healthy relationship cultivates stability between the parent and child, allowing the child to attach more securely to the parent, which facilitates placement permanency and positive outcomes (Mountjoy & Vanlandingham, 2015). A child's ability to form attachments within the parent-child relationship is often a predictor of future internalizing and externalizing behaviors (Beijersbergen, Juffer, Bakermans-Kranenburg, & van Ijzendoorn, 2012).

> Attachment is the enduring emotional closeness that binds families, to protect children and prepare them for independence and parenthood. … Early attachment establishes preconceptions of the value, reliability, safety and use of relationships, with lifelong implications for the extent of emotional self-sufficiency, and for behavior in relationships. (Rees, 2008, p. 219)

Adoptive children are overrepresented in mental health and special needs services, in part due to suffering from low self-esteem, exhibiting a lack of academic achievement, and developing behavioral problems with some presenting as psychiatric disorders (Juffer & van Ijzendoorn, 2005; van Ijzendoorn & Juffer, 2006; van Ijzendoorn, Juffer, & Poelhuis, 2005). However, research has shown that the majority of adoptees are well adjusted (Hjern, Lindblad, & Vinnerljung, 2002; Stams, Juffer, Rispens, & Hoksbergen, 2000; Tieman, van der Ende, & Verhulst, 2005, 2006; Verhulst, Althaus, & Versluis-den Bieman, 1990). The adoptive family experience resides in a complex context limiting the value and impact of analysis. As discussed above, the research findings are often contradictory and reveal small effect sizes, which weakens the conclusions and does not provide a consistent clear path to success for practitioners as well as adoptive families. The relational well-being subset of the NSAP could be the tool used to delineate success within the adoptive family. The use of item response theory to analyze the relational well-being subset of the NSAP measure is ideally suited to establish a psychometrically stable benchmark of the assumed relational well-being construct within the survey.

Item Response Theory

Item response theory (IRT) describes the interactions between persons and test items (Reckase, 2009). Dimensional structure, model fit, item fit, reliability, and validity are used to characterize the model fit through the person/item interactions within IRT. There are at least three advantages of IRT over classical test theory (CTT): (1) diagnostic indices are available to assess the data fit to the model at the item-, person-, and model-level; (2) statistical tools aid in establishing the optimal categorization of rating scale structures; and (3) conditional standard errors support the precision of estimates for the examination of varied levels of person position (Fan, 1998; Sharkness, 2014; Sharkness & DeAngelo, 2011). Depending upon the type of measure, the person's ability indicator describes the individual's amount of agreement or the amount of a latent trait (Bond & Fox, 2007). The probability of success or a person's ability to endorse an item is calculated within the context of the person's ability and the item's difficulty (Bond & Fox, 2007).

The assumptions of unidimensional IRT include monotonicity, unidimensionality, and local independence. Raw scores collected from the measure have been proven algebraically to be sufficient to determine a person's ability. Monotonicity represents the relationship between the latent trait and responses (S-curve, Figure 1). Next, IRT assumes that the measure represents only one construct, so is unidimensional. Finally, local independence assumes that the items are not dependent on each other. Georg Rasch (Bond & Fox, 2007) developed the first dichotomous IRT model, which utilized a logarithmic transformation of ordinal data into interval data. The original Rasch model was extended from a dichotomous model to models with the capability of transforming polytomous response scales as well, called the Rasch rating scale model (RSM) (Andrich, 1978). Construct validity is established through a Rasch analysis via adequate model fit showing unidimensionality and measurement invariance, an ordered item and person continuum that reveals enough variation from difficult to easy items (probability of endorsement) and knowledgeable to unknowledgeable persons (probabilities of success), and item characteristic curves (ICC) not crossing (i.e., similar slopes). Figure 1 provides a view of the relationship between the trait and probability of item response.

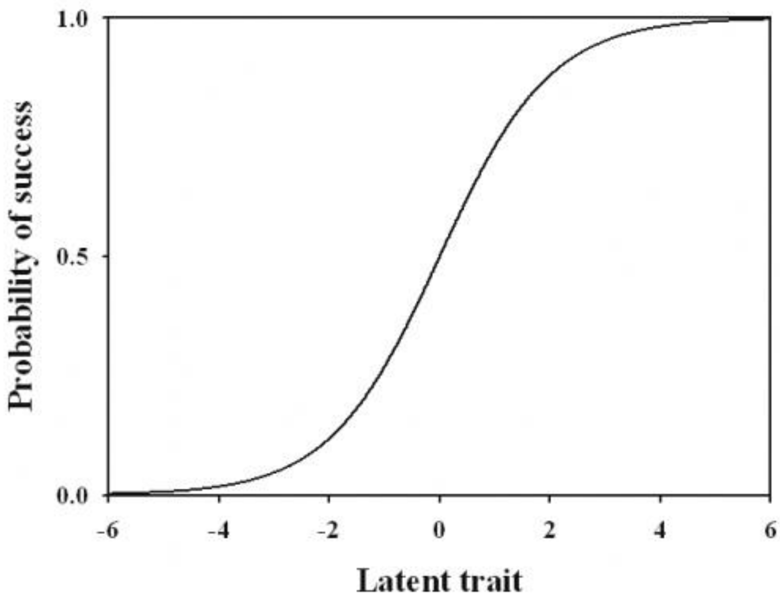


Figure 1. Item characteristic curve showing the relationship between location on the latent trait and the probability of answering the item correctly.

A person with an ability of 0.0 on the latent trait has a probability of .5 of answering the item correctly or endorsing the statement. The probability of endorsing a statement increases as a person’s ability, or position on the trait, increases. Crossing ICCs for different items reveals that the difficulty characteristic is no longer isolated to the item itself but to the item and the person’s ability, which eliminates a strength of a Rasch analysis. For polytomous response items, category response curves (CRC) are used to compare a person's ability and the probability of a correct response or item endorsement, on each of the scale options. Figure 2 below shows that with a person ability of 1 there is a 10% chance of answering response 2, a 30% chance of selecting 3, and an 10% chance of selecting response 4.

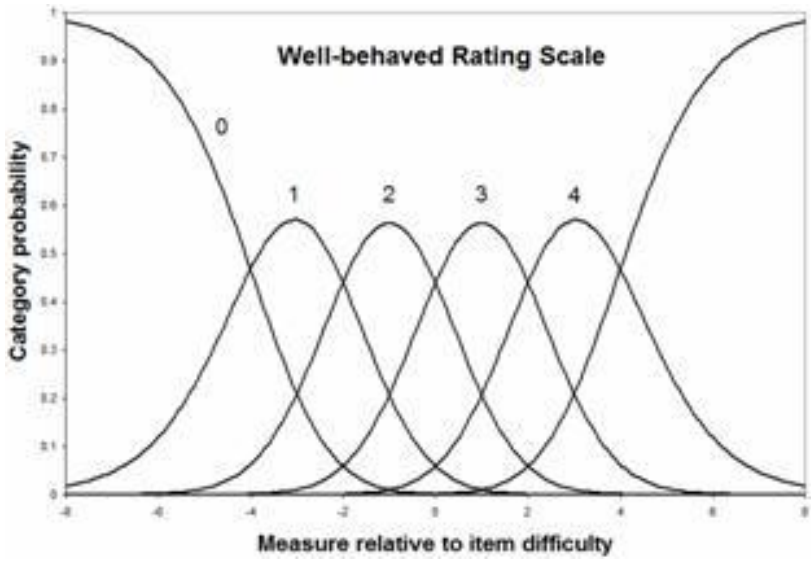


Figure 2. Category probability curves for a polytomous item.

In the partial credit model (PCM) the item format and categories distances vary for each item. The PCM is appropriate when items have different response categories in either wording or number. For example, if one item has yes-no response options while another item has strongly agree to strongly disagree response options, a PCM would be used in measure development. Rasch PCMs test fit through a series of fit indices: person fit, item fit, dimensionality, and differential item function (DIF). Model fit within Rasch models utilize mean squares with an expectation of 1.0, with a range between 0.0 and infinity. A mean square fit of 1.0 indicates the data fit the model perfectly. Underfit items or persons have values greater than 1.0, which identifies excessive noise within the data. Overfit items or persons have mean squares fit values less than 1.0, which indicates the possibility of overlapping or muted content. Identifying excessive noise within the data is typically considered more valuable than concerns of potential overlapping content. Misfitting items are seen as not fitting the construct. Fit statistics transformed into z-standardized statistics (zstd = 0 and MS = 1.0) can be utilized when the sample size being tested is small or if there are few items in the measure. Accepted misfit z-standardized statistic cutoffs with samples of 30-300 subjects are underfit, zstd >2.0, indicating too much variation and overfit, zstd<-2.0, indicating too little variation (Bond & Fox, 2007).

Person and item fit are measured in two ways in the Rasch model. Infit is an index calculated by weighting the measure by the distance between person and item location, while the outfit index is an unweighted measure. These indices are transformed chi-square statistics. Satisfactory infit and outfit values for items with polytomous response scales have a range of 0.6-1.4 (Bond & Fox, 2007; Rahayah Ariffin, Omar, Isa, & Sharif, 2010). Items with more variability than expected are found to have values above 1.4, while those items with less variability than expected have mean square fit values below 0.6. Item fit describes the functioning of the items in the context of the model. Items that fit a single construct, forming a continuum, and are logical within the model content are classified as having good item fit. Poorly fitting items tend to be too complex or difficult in relation to the whole scale, or may not be measuring the single construct being examined by the instrument (Rahayah Ariffin et al., 2010). The item information function (IIF: Figure 3) provides item level information, which allows the researcher to tune the measure's items.

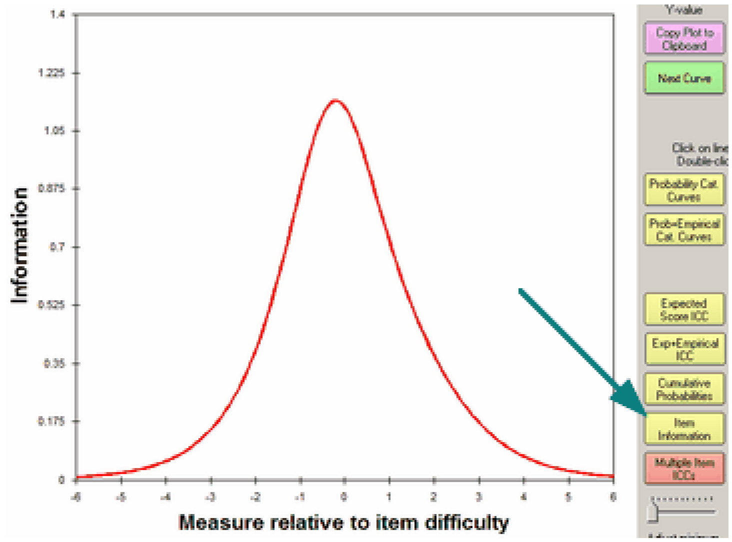


Figure 3. Item information function for item level information, by J. Linacre, 2012.

The IIF indicates the precision and reliability of the responses to an item relative to person ability. According to Figure 3 above, individuals with an ability between -2 and 2 logits of the item location are in the optimal range for this item. Individuals with abilities outside the optimal range will tend to produce less consistent responses. Items that correlate highly, at a level of 0.9 or higher, as determined by Mokken scaling (a specific IRT model), could be trimmed, since these items are considered to be measuring the same thing. The assumed independence of the items within IRT support an additive component to building the measure. With the assistance of fit indices, the test information function (TIF) guides the development of the most effective measure with the least number of items.

Person fit indices reveal the consistency of the individual's responses. The better the person fit, the more consistently the individual's response matches the Rasch model expectations. The person reliability indicator shows if the measure is sensitive enough to distinguish between high and low levels of performers for the particular sample being tested (Bond & Fox, 2007).

Additionally, Rasch models graph person-item location by the positions of items and persons in relation to each other—the Wright map. This graph is useful in examining the degree to which items and persons match. The gaps between items can be audited to determine where on the continuum of difficulty items need to be added or removed, allowing for a more complete and parsimonious item continuum (Bond & Fox, 2007).

The rating scales or categories of each item can be calibrated by collapsing the unnecessary or rarely used points on the item-scale. Probability curves provide a visual diagnostic tool for rating scale function and support the calibration process in establishing uniformly spaced rating scales or ordered categories (Royal, Ellis, Ensslen, & Homan, 2010). Combining or removing rarely used scale points can improve the measurement quality and fit of the model but this type of adjustment can also decrease fit as well. Polytomous items are more complicated and susceptible to needing category calibration. The structural calibration of response scale categories is expected to progress in order; otherwise, category disorder is observed. In partial credit models, the item format and categories distances vary for each item, as necessary, while the rating scale model constrains all item categories to the same relative distances (Bond & Fox, 2007).

Measure invariance supports the “sample free” assumption, which allows the item estimates to be considered independent of the distribution of persons responding to the items. Established invariance allows the researcher to use the measure as a consistent measure of the perceived construct regardless of the person’s ability, time the measure is administered, or group characteristics. The item and person fit, the correlation of item logit position by groups, and the differential item functioning (DIF) test are all indices for examining measure invariance within a Rasch analysis. DIF tracks the change in variable meaning by examining the item location with respect to different groups (Bond & Fox, 2007). Rasch software calculates logit position by group, while dividing the difference in position by the combined standard error to generate a significance test of differences in group logit positions.

Items with a logit difference greater than or equal to 0.50 at $p < 0.01$ between groups are considered as showing signs of group variance. The meaning of the variable is violated if DIF is found for the items, which evidences between group differences or misunderstanding of items (Bahraini, 2008; C.C. Chang et al., 2015; Cheng, Wang, & Ho, 2009). DIF can be affected by both the effect size and the group size. If invariance is not achieved, then an instrument assesses a construct that is understood differently by different groups, and yields scores that cannot be compared across groups.

Unidimensionality is assumed in Rasch modeling, which means that the collection of items within the instrument are expected to represent a single construct, see seen in Figure 4. The example assumes that the (thirty in the example) items being analyzed load on the latent construct "well-being." Items within an instrument that fit poorly are removed to improve the unidimensionality of the instrument. Evidence of a single construct within a Rasch model indicates validity within an IRT model (Yu, Popp, Digangi, & Jannasch-Pennell, 2007). Unidimensionality is supported if the explained variance of the model is > 40% and the eigenvalue for the first contrast is < 2.0 (Bond &

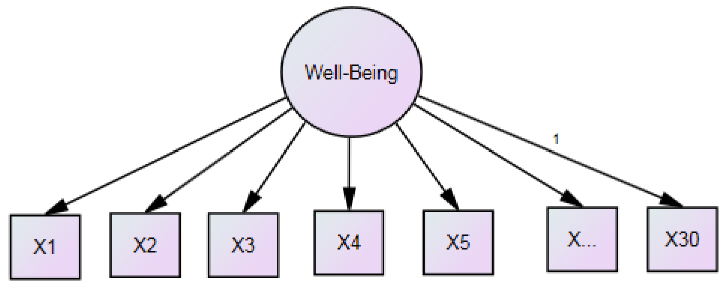


*Figure 4***.** Rasch unidimensional model example.

Fox, 2007; Linacre, 2012). These indicators help the researcher to determine if a potential second dimension is due to more than just chance. Eignevalues of 1.4 are accepted as the threshold of random noise (Smith & Miao, 1994). At times, a potential second dimension is considered, despite dimensionality indices not reaching the thresholds of the empirical indicators, when a review of item content conceptually supports a second dimension. Thus, measure design intent and content review are privileged beyond simple review of numerical indices in determining if a second dimension is sought. For example, the Brief Symptom Inventory-18 (Derogatis, 2001) was designed to assess three dimensions, and thus three-dimensional models are tested whether or not numerical indices indicate adequate fit to a unidimensional model.

Once the unidimensionality of a measure has been brought into question, two different Rasch approaches can be used to examine the dimensionality of the data. The first approach, identified as the consecutive approach, uses an iterative process to identify potential dimensions. The consecutive approach begins similarly to the unidimensional approach except the analysis is repeated with the removed misfitting items in order to identify emerging dimensions. As items are removed from the model/dimension, the misfitting items are re-pooled and evaluated to find if an additional stable dimension is present. When analyzing a pool of items for the first time an eigenvalue of >2.0 may indicate a subsidiary dimension within the subsequent misfitting items. The same criteria of explained variance of > 40%, eigenvalue of the first contrast < 2.0, and item MS infit of 1.4 to 0.6 are used (Bond & Fox, 2007; Linacre, 2012). The process is repeated until there is no collection of items meeting the cutoff criteria. Figure 5 provides an example of five latent constructs found through the consecutive approach with the specific items found loading on each construct. The consecutive approach produces individual dimension estimates and standard errors but loses the potential interaction between the dimensions due to separating the dimensions.

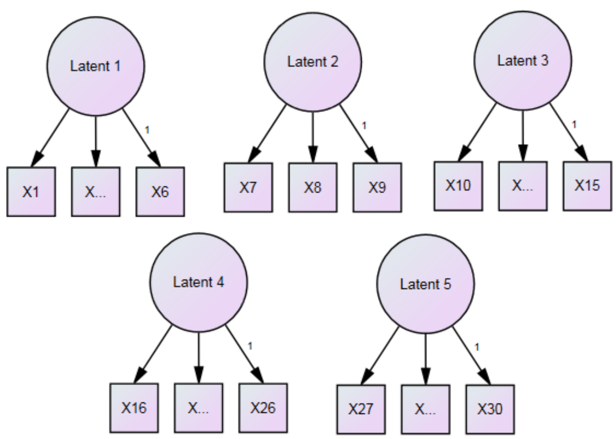


Figure 5. Rasch consecutive approach example.

Reliability for the consecutive approach may be lower than for a unidimensional approach, since the standard error estimates are larger in the consecutive approach (Briggs & Wilson, 2003). The third approach is multidimensional, which is used to confirm suspected multidimensional models, and is an enhancement between the unidimensional approach and the consecutive approach that utilizes dimensional correlations, as seen in Figure 6.

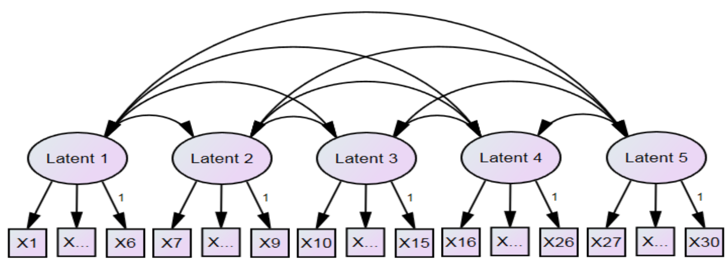


Figure 6. Rasch multidimensional model example.

The multidimensional approach affects reliability through the use of the inter-relationships between the dimensions and reduced standard error estimates, unlike the consecutive approach (Allen & Wilson, 2006; Briggs & Wilson, 2003). The multidimensional approach can be used to confirm dimensionality suspected by the researcher or test dimensionality found through EFA, PCA, CFA, or a consecutive approach while using an item response theory model.

Multidimensional Rasch is an extension of unidimensional IRT, when the measure assesses multiple constructs. Multidimensional Rasch accepts the complexity of the data, while idealizing reality through the approximation of person ability and item difficulty (Ackerman, Gierl, & Walker, 2003; Briggs & Wilson, 2003; Reckase, 2009). The multidimensional random coefficients multinomial logit model (MRCMLM) is a flexible model that allows for nonzero correlations between latent constructs and fits a variety of data. The unidimensional approach becomes inadequate to explain the data when data are determined to be multidimensional (Cheng et al., 2009). MRCMLM is used to increase the validity of multidimensional measures that contain dichotomous and/or polytomous data by estimating model fit, person fit, and item fit (Allen & Wilson, 2006; Rost & Carstensen, 2002).

Once the model has been evaluated through the three approaches: unidimensional, consecutive, and multidimensional, the models can be compared via the estimated model fit, deviance, reliability, correlations, the likelihood ratio statistic, G^2–similar to $\chi 2$ with degrees of freedom matching the parameter count difference between models--and Akaike's Information Criterion (AIC) for relative model fit (Akaike, 1974; Allen & Wilson, 2006). If a statistically significant difference in deviance is found between models, then the difference in deviance is large enough to support the more complex model as a better fit to the data. A nonsignificant difference in deviance supports the more parsimonious model. In addition, AIC is used to compare model fit. The model with lowest AIC value would indicate the best model fit between the unidimensional, consecutive approach, or multidimensional approach provided the same items were used (Allen & Wilson, 2006; Purya Baghaei, 2013; Briggs & Wilson, 2003; H. L. Chang & Shih, 2012). Akaike (1985) suggests that the AIC can be used, in principle, to compare nonnested models. A nested model is defined as a smaller or simpler model found within a larger or more complex model for comparison. These models are compared using likelihood ratio tests, such a G^2, or identifying which model explains more of the variance. Nonnested models are defined as models that cannot be derived from one another through parametric restriction or limiting. Nonnested models may describing different segments of the variance within the data. Dimensions within the nonnested models can include interrelationships between the dimensions or the interrelationships may be absent.

Studies Comparing Unidimensional, Consecutive, and Multidimensional Models.

A psychometric study, examining the difference between multidimensional models and unidimensional models on scales of willingness to communicate in a foreign language, found that low to moderate correlations between dimensions supported multidimensionality. A Rasch multidimensional analysis was conducted to evaluate the best fitting model. It was determined that the 3-dimensional model was the best fitting (Purya Baghaei, 2013). An analysis by Allen and Wilson (2006) of health behavior and health education research stated that the composite (unidimensional), consecutive, and multidimensional approaches each have their advantages. For traditional models, the composite approach is most parsimonious and direct way to model the data. The consecutive approach allows the researcher to examine the subscales of the multidimensional model. The multidimensional approach provides a complex representation of the data and reduces the overestimation of measurement error from the consecutive approach, while adding insight into the relationships between the dimensions. Both the unidimensional model and the multidimensional models allow for simpler and more direct interpretation than the consecutive approach.

A comparative analysis of a student achievement measure using the composite, consecutive, and multidimensional approaches found that the multidimensional approach provided the best model fit (Briggs & Wilson, 2003). The consecutive approach explored the measurement estimates of the subscales through use of multiple unidimensional models. Person estimates were unnecessarily larger and the reliability estimates were smaller when using the consecutive approach. The reliability estimates produced by the multidimensional model were closer to an overarching unidimensional model. A concern of this research was that simplifying to a unidimensional model misrepresents the person ability, especially when the examined dimensions in the multidimensional model have a low correlation (Briggs & Wilson, 2003).

Wiberg (2012) examined the impact of a model for a college admissions test using unidimensional, multidimensional, and consecutive approaches. The analysis found that the multidimensional model showed better fit than the unidimensional model, which resulted in the poorest fit. There was a concern when using the consecutive approach for the multidimensional subset when too few items, less than 20 items, were present. There were small differences between the consecutive approach and the multidimensional approach in favor of the multidimensional approach. This author supported the idea that the consecutive approach led to easier interpretation of the dimensions than the multidimensional approach because the item subsets were isolated around the topic within the college admission test and so could be interpreted independently (Wiberg, 2012).

The choice to use multidimensional IRT for this project was due to prior analyses that utilized classical test theory in the form of principal components analysis and confirmatory factor analysis (Park et al., 2013). The results of these analyses indicated the NSAP has a multidimensional structure, though with many of the 49 original items removed. The large amount of procedurally planned missing data in the NSAP dataset may have contributed to the decision in earlier analyses to use only a few of the original 49 items. IRT analyses are better suited for dealing with missing data than classical test theory, while providing information on item fit and difficulty, person fit and level, and an evaluation of the item response scale effectiveness and item targeting. Since the IRT analysis examines each item individually, while classical test theory evaluates the items within the whole test together, the impact of missing data is seen on the items with missing data via larger standard errors and less so on the entire measure. This does not suggest that missing data have no impact on an IRT analysis, rather that the impact is lessened by the technique. Additionally, item data that are planned to be missing have less of an impact on the analysis than items skipped by the participant. Items skipped could be random or skipped due to an item-related rationale. The items purposely skipped may result in an over-or underestimation of the item and person fit (Bolsinova & Maris, 2016; DeMars, 2002).

The NSAP dataset with approximately 2089 cases was randomly split into two subsets with the goal of providing balanced samples for item response analysis. Unidimensional and multidimensional approaches were utilized on the first half of the data to determine the best fitting model. The best fitting model from the findings of the initial half of the data was applied to the second half of the data for replication. The Rasch analysis of the NSAP items clarifies the dimensionality of the well-being subset as well as enhances the usefulness of these items for future research, while providing insight into potential areas of improvement for the measure.

The traditional unidimensional approach of Rasch supports a simple direct use of a measure or survey, which allows for clearer interpretation. The consecutive approach explores the dimensional potential by removal of the best fitting items and then repeating the unidimensional analysis process. By examining the items remaining once the best fitting items are removed, the researcher can search for other potential dimensions. This process continues until all items are accounted for within a new dimension or are eliminated from consideration. The multidimensional approach includes the interactions of predetermined dimensions within a measure to determine model fit. This approach is used for complicated models, where the researcher is attempting to address constructs with overlapping characteristics. Comparing these approaches with a single measure allows the researcher unique insight into each process as well as determining the best model fit for the exemplar data.

The end goal of measure development is appraisal of validity. Validity, from a Rasch perspective, comprises reasonable item and person fit and appropriate progression of item position that reflects understanding of the construct. However, validity is more widely inclusive of evidence related to the utility of a measure in prediction of a desired outcome and in convergence with measures thought to be related to the measure under study.

Additionally, comparing the impact of various attributes upon outcomes across dimensions and models allowed for increased legitimacy of the findings. NSAP collected a variety of demographic characteristics that have been used to evaluate the consistency of the dimensions between models for validation. Three of these variables used in this study describe potential characteristics of adoptive families. The characteristics used were Adoptive Family with or without Biological Children, Child Lived with Birth Family, and Adoptive Parent/s and Child of Differing Races.

Chapter Two: Methods

This chapter delineates how this research study was conducted by providing a description of the dataset, sample population, and variables used. Sections included describe how the survey was developed and how the data were collected, the data splitting procedures for this study, and the analytic methods used for each research question. The analyses were performed to determine the dimensionality of the parent-child relational well-being subset of the NSAP data.

Survey Development and Data Collection Procedures

The National Survey of Adoptive Parents was conducted from April 2007 to July 2008 by the Centers for Disease Control and Prevention's National Center of Health Statistics (NSCH) as an add-on to the 2007 National Survey of Children's Health to establish national estimates of adoptive children and their families' well-being, health, and other characteristics. A survey with a focus on the entire adoptive community in the United States had not previously been fielded. In 2005, the Assistant Secretary for Planning and Evaluation (ASPE) requested that both the Urban Institute and NORC of the University of Chicago develop an instrument for the National Survey of Adoptive Families. The both groups reviewed the adoptive research literature. General topics of interest were established and research findings and existing adoption items were organized into topic areas of interest. Then, the topic areas and items were categorized based upon the perceived level of importance. An initial survey was sent to ASPE for review, which generated suggestions regarding new item wording and missed topic areas (Bramlett, Brooks, et al., 2010).

At the next phase of survey development, cognitive interviews were conducted with seven adoptive parents (five foster-to-adopt parents, one private domestic parent, and one inter-country parent). The survey developers explored how the items were perceived by the adoptive parents and the researchers made adjustments as needed. In the final step of the survey development, eight adoptive parents (two foster-to-adopt parents, three private domestic parents, and three inter-country parents) participated in a pretest of the final survey draft in order to determine time needed to complete the survey (Bramlett et al., 2010). It should be noted here that the survey was not developed to reflect any particular theory.

Using a random-digit dialing method and a module of the State and Local Area Integrated Telephone Survey (SLAITS), the NSAP obtained a nationally representative sample of adopted children under 18 and interviewed in English the adoptive mother or adoptive father of each selected child. Children who lived with a biological parent were excluded from the sample. The average phone interview for the entire survey lasted 30:46 minutes (median time = 29:24 minutes) with the well-being subset lasting 3:30 minutes on average (median time = 3:18 minutes) (Bramlett et al., 2010; Centers for Disease Control and Prevention, 2009). NSAP was the first study to use a nationally representative sample representing all types of adoption, inter-country, foster-to-adopt-, or private domestic in the United States. Sampling weights, cluster weights, and strata weights were applied to the data to represent the national population of adoptive parents. Participation in the NSAP survey was voluntary and confidential. The survey had a 34.6% response rate with a 74.4% completion rate of participants (Bramlett et al., 2010).

Due to protocol skip logic, not all participants were asked all of the questions within the well-being section of the NSAP. Participant groups were determined by age of the child at the time of survey into the adoptive home. The groups were identified as children under 6 months old, children 6 months old or older but younger than 1 year, children 1-year-old or older but younger than 5 years old, children 5 years old or older but younger than 13 years old, and children 13 years old or older. The NSAP data were available on the Centers for Disease Control and Prevention website at https://www.cdc.gov/nchs/slaits/nsap.htm. Through the use of SAS version 9.3, the data were downloaded and converted to an excel file for input into Rasch software.

Participants

The NSAP sample included 2,089 parents whose adoptive child was 17 years or younger and was still living in the parents' home. The majority of respondents were adoptive mothers (79%, n = 1,651), about a fifth were adoptive fathers (20.2%, n = 423), and .8% were not clearly identified (Bramlett et al., 2010; Park et al., 2013).

Human Subjects Protection

The original project conducted by Centers for Disease Control and Prevention's National Center of Health Statistics received human subjects approval through National Survey of Children's Health Research Ethics Review Board, December 2006, and the University of Chicago Institutional Review Board, November 2006, in compliance with

Health and Human Services regulations (45 C.F.R. 46) (Bramlett, Foster, et al., 2010). For this research, an exemption request was submitted to Institutional Review Board at the University of Denver for analysis to be conducted on the public de-identified NSAP dataset. The use of the NSAP data was approved by the Institutional Review Board on May 17, 2017 with an expiration of May 16, 2020, project number 1066213-1. The primary data used from the NSAP dataset were the items from the Well-Being subset of the NSAP, see Appendix A. Other data points used for this study were collected from the NSAP screener section, characteristics section, or established by data collection procedures. The participant ID was generated through the data collection procedures and only used in the beginning of this study. A unique ID number was generated for this research to simplify the ID use. Adoption type data, used in this research, were collected during the screening portion of this survey. The age of the child during the administration of the survey, the racial difference between the parent and child, if the adoptive parent had biological child/ren, and if the selected child lived with their birth family at any time were all collected in the characteristics portion of the NSAP survey.

Data Randomization and Splitting Technique

The dataset was split into two balanced sets based on data collection protocols established during the during the initial survey administration between April 2007 and July 2008. The balanced datasets allowed for a comparison of the results due to the similarities between the datasets as well as to the original complete dataset. The five protocol groups were determined by the age of the child at the time of the initial survey and group counts are shown in Table 1.

Table 1

National Survey of Adoptive Parents Well-Being Data Collection Groups

Collection Groups	Total
Child < 6 months old (Group 1)	5
Child > 6 months old & < 1 year old (Group 2)	206
Child > 1 year old & < 5 years old (Group 3)	151
Child > 5 years old & < 13 years old (Group 4)	958
Child > 13 years old (Group 5)	769
Total	2089

Note. Collection Groups were established by the data collection protocols of the National Survey of Adoptive Parents, which were based upon the child's age.

In addition to balancing the datasets by collection group, the two datasets were balanced between adoption types: international participants, foster-to-adopt participants, and domestic-private participants. Collection Group 1, adoptive families with children younger than 6 months old, were excluded from this analysis, since there are only five responses and all were within the domestic-private adoption type. Once age group and adoption type separated the data, the participants were placed randomly into the two split halves for analysis. Each set (Table 2) contained 1,040 participants with 206 responses from Groups 1 and 2, 150 responses from Group 3, 956 responses from Group 4, and 768 responses from Group 5.

Table 2

NSAP Collection Groups by Adoption Type and Analysis Type

Adoption Type	Exploratory	Confirmatory

International	272	272
Child > 6 months old & < 1 year old (Group 2)	46	46
Child > 1 year old & < 5 years old (Group 3)	29	29
Child > 5 years old & < 13 years old (Group 4)	137	137
Child > 13 years old (Group 5)	60	60
Foster-to-adopt	381	381
Child > 6 months old & < 1 year old (Group 2)	22	22
Child > 1 year old & < 5 years old (Group 3)	26	26
Child > 5 years old & < 13 years old (Group 4)	168	168
Child > 13 years old (Group 5)	165	165
Domestic -Private	387	387
Child > 6 months old & < 1 year old (Group 2)	35	35
Child > 1 year old & < 5 years old (Group 3)	20	20
Child > 5 years old & < 13 years old (Group 4)	173	173
Child > 13 years old (Group 5)	159	159
Total	1040	1040

Note. Collection Groups were the groups each participating parent was placed in based upon the age of the child. These collection groups impacted the questions received by the participants. Adoption Type identifies the adoption category the adoptive parent and child experienced. The data were further split into Analysis Types for this particular research project. The two types were identified as exploratory and confirmatory, depending on whether the responses were placed into the initial phase of the research or were used to confirm the findings in the second phase.

Once the datasets were created and balanced by group along with the adoption type, the overall group count of the analyzed responses were compiled as shown in Table 3. The resulting dataset contained 2,080 participant responses with nine of the original responses unused for this project.

Table 3

NSAP Collection Groups Retained for Study

Collection Groups	Total
Child > 6 months old & < 1 year old (Group 2)	206
Child > 1 year old & < 5 years old (Group 3)	150
Child > 5 years old & < 13 years old (Group 4)	956
Child > 13 years old (Group 5)	768
Total	2080

Note. Collection Group counts retained from the balancing procedures.

Software

Demographic analysis was done through use of IBM® SPSS® statistical software (version 22), construct/dimensionality charts drawn through IBM® SPSS® Amos (version 24); and Winsteps 3.92.1 and ConQuest (version 4) of the Australian Council for Educational Research for the Rasch modeling. Additional information was obtained via use of Rasch software, Winsteps (Linacre, 2017). Winsteps utilizes joint maximum likelihood estimation (JMLE) algorithm to establish parameter estimates. This estimation algorithm calculated the item and person estimates, while accounting for the scale of the item. ConQuest utilizes the marginal maximum likelihood estimation (MMLE) algorithm to calculate parameter estimates. MMLE calculated the item and person estimates and then took the item scale into account (Linacre, 2012).

Analytic Strategy

Research question one.

Did the data from the parent-child well-being subsection found in National Survey of Adoptive Parents (2007), NSAP, support a unidimensional or multidimensional structure when using a Rasch partial credit model for analysis?

a. Were the psychometric properties of model fit, item fit, and reliability more suitable for the NSAP data within a unidimensional, consecutive, or multidimensional model?
b. How did item and person logit positions differ between the unidimensional, consecutive, and multi-dimensional findings?
c. Did item and person indicators of position differ between software (Winsteps and ConQuest) when using the unidimensional and consecutive approaches?
d. Which approach yielded a better model fit for the well-being subsection of NSAP?

With an initial assumption of unidimensionality, an exploration of the dimensionality of the first randomly generated half of the parent-child well-being subsection found in NSAP data was completed through the use of a Rasch analysis. Unidimensionality was supported when the explained variance was greater than or equal to 40%, the first contrast eigenvalue was less than 2.0, and the item infit and outfit indicators were within acceptable ranges. See Appendix G for definitions of indices used in the analyses. Next, the consecutive approach process with the first half of the split dataset used in research question one was conducted. For this study the consecutive approach was utilized as an exploratory tool to determine possible dimensions within the data. Dimensions were isolated through the consecutive approach until all fitting items were included. Once the number of dimensions and the items reflecting those dimensions were established through the consecutive approach, analysis ceased. The established dimensions were then analyzed through multidimensional IRT for model fit. Fit indices (deviance, AIC) of the unidimensional, the multi-dimensional of the consecutive solution, and the multi-dimensional model were then compared. Model estimate parameters were compared between the two software types to examine the similarities and differences since different estimation algorithms are used. Once the analysis was completed, the optimal model was selected for examination by the second half of the original split dataset. Item fit, reliability, and model fit via deviance and AIC were all considered in selection of the optimal model.

Cutoff criteria were used to identify items and persons to be retained in the analysis as follows:

1. Item infit and outfit mean square cutoff criteria range between 0.6-1.4 (Bond & Fox, 2007).
2. Person fit was examined and person records sparingly deleted (when the infit mean square was over 9.0).
3. Compare likelihood ratio test G^2 at $p < 0.05$ to identify significantly better fit between models (Allen & Wilson, 2006).
4. Identify model with lowest AIC (Allen & Wilson, 2006).

Research question two.

Were scale response categories appropriately employed for each of the utilized items?

Additional analysis was conducted as part of determining the best fitting model from the first split of the data. The response scale was examined to determine the need for scale recalibration. Items with a response category mean square outfit of greater than 2.0, an underutilized (< 10 respondents) category, and/or items with categories disordered beyond a standard error boundary were identified as needing to be adjusted. Those categories within the item identified as needing adjustment were collapsed, which resulted in a decrease in the number of categories within the rating scale. The final calibrated model was compared to the model of research question one to determine if the scale adjustments improved model fit.

Research question three.

Once the dimensionality had been established and the item categories had been determined, did respondents for different adoption types (inter-country, domestic-private, and foster-to-adopt) interpret the items differently as observed through differential item functioning (DIF)? How did the differential item functioning results for adoption type compare between the unidimensional, the consecutive, and the multi-dimensional approach?

Once the dimensionality and the best fitting model were established, invariance was examined. The group classifications were the adoption types of international adoption, foster-to-adopt, and private-domestic adoption. Each of the items within the identified dimension/s were examined. If a logit difference greater than or equal to 0.50 at $p < 0.01$

was found, the item was considered as showing signs of group variance and respondents for different adoption types interpreting the item differently (Cheng et al., 2009).

Research question four.

How did the person logits and item difficulty compare for each dimension found within models from the unidimensional, consecutive, and multidimensional approaches compare across models?

a. Did the person logit positions correlate across dimensions/models and software packages?
b. Using a cluster of three independent variables, Adoptive Family with or without Biological Children, Child Lived with Birth Family, and Adoptive Parent/s and Child of Differing Races, as predictors in a regression analysis were the R^2 values comparable between the models and the software?
c. Using the same cluster of family characteristic variables as independent variable in a canonical correlation analysis were the canonical R^2 values comparable between the models and the software?
d. How did the item difficulty compare across models examined?

Person logit position was correlated across the models and software, Winsteps and ConQuest, to determine consistency. It was expected that models representing the same dimensions would correlate highly despite different software being used. Dimensional correlation estimates were examined to provide insight into the decision regarding impact of analysis approach (unidimensional, consecutive, multidimensional).

Next, person logit position was used as the dependent variable(s). For models with a single outcome, the unidimensional model, a linear multiple regression was conducted with three predictor variables: Adoptive Family with or without Biological Children, Child Lived with Birth Family, and Adoptive Parent/s and Child of Differing Races. These independent variables reflect family traits with perceived impact on the parent child relationship (Bramlett, Foster, et al., 2010). These data were coded as binary variables. The R^2 indicates the amount of variance predicted by the independent variable on the dependent variable, person logit position. Models and dimensions, with similar items, that have larger R^2 values indicate a better solution. For models with multivariate outcomes, such as the multidimensional models and the consecutive models, correlation with family characteristics led to the use of canonical correlation analysis. From these analyses the canonical R^2 was calculated for comparison with the R^2 from the regression.

Item difficulty was compared across each dimension/model and software used. The dimensions developed through the consecutive approach were combined and correlated with the unidimensional, the 3-dimensional, and the 2-dimensional models in order to give a complete view of the item difficulty differences. Item consistency was also evaluated through this technique.

Research question five.

With the best fitting model selected, were the dimensionality and model fit replicated through the use of a second half of the dataset for cross validation? Were the item fit, DIF, and validation measures comparable across the two halves of the dataset?

The optimal model was examined using the second half of the split dataset as confirmation of the initial findings. The established dimensions were analyzed through ConQuest for model fit. Infit, outfit, deviance, and AIC were used to confirm consistency with the previous findings.

Chapter Three: Results

This chapter reports the results, based upon questions identified in Chapter One, for unidimensional, consecutive, and multidimensional Rasch analyses of the NSAP well-being data.

Research Question One

Did the data from the parent-child well-being subsection found in National Survey of Adoptive Parents (2007), NSAP, support a unidimensional or multidimensional structure when using a Rasch partial credit model for analysis?

a. Were the psychometric properties of model fit, item fit, and reliability more suitable for the NSAP data within a unidimensional model or multidimensional model?

b. How did item and person logit positions differ between the unidimensional and multi-dimensional findings?

c. Did item and person indicators of position differ between software (Winsteps and ConQuest) when using the unidimensional and consecutive approaches?

d. Which approach yielded a better model fit for the well-being subsection of NSAP?

Unidimensional model

The well-being section of the NSAP survey contained 39 items and all were considered for the unidimensional model utilizing Rasch partial credit analysis through Winsteps. During the initial execution of the analysis, 20 items were dropped due to a lack of data, with ≥ 99% missing data, while 19 items retained. The removed 20 items were only administered to participants that met very particular characteristics. None of the 20 items with missing data were used in any further analyses.

The initial variance explained was 58.0% with a first contrast eigenvalue of 1.78 and a first contrast percentage of 3.9%. The person model separation was 1.94 with a person reliability of 0.79 and Cronbach's Alpha of 0.86. The item model separation was 11.06 with a model reliability of 0.99. Items with an infit mean square or an outfit mean square over 1.4 were considered to be misfitting. A standardized Z (ZSTD) score was included to show the likelihood of the data fitting the Rasch model by chance. The ZSTD scores ≥ 3.0 suggest unpredictability in the responses, while ZSTD scores ≤ -2.0 may indicate too much predictability in the responses (Linacre, 2002). Twenty-one respondents were removed from this analysis due to person misfit with person infit mean square of over 9.0. Through an iterative process seven items were removed from consideration due to misfit mean squares and misfit ZSTD, see Table 4.

Table 4
Misfitting Items found during Unidimensional Analysis

	Infit		Outfit		
Item removed	MS	ZSTD	MS	ZSTD	Iteration

W17AR	3.59	7.6	9.90	9.9	2
W17	1.47	4.5	9.90	9.9	2
W18	1.65	3.3	1.92	3.5	3
W16	1.11	0.8	3.40	3.9	3
W9	1.51	6.6	1.55	6.6	4
W17B	0.99	0.0	1.60	1.1	4
W5	1.50	8.7	1.48	8.1	5

Note. MS = mean square and ZSTD = Z-standardized.

The final model, identified as the unidimensional model, included 12 items with a mean MS infit of 1.00 and a mean MS outfit of 0.97. The item infit mean squares ranged from 0.77 to 1.34 and the item outfit mean squares ranged from 0.65 to 1.38. Item fit statistics can be found in Table 5. These 12 items resulted in 56.4% of the total variance explained by the measure with 37.0% of the explained variance due to the person and 19.3% of the variance due to the items. The person model separation was 1.61 with a person reliability of 0.72 and Cronbach's Alpha of 0.90. The item model separation was 15.81 with a model reliability of 1.00. During the trimming process, items W14 and W4R were recalibrated based on inversions in Andrich Thresholds, so the categories were better aligned with the data (see following section). The result of the remaining 12 items was labeled "unidimensional."

Table 5
Item Fit Information for Final 12 Items in Unidimensional Model

	Infit		Outfit	
Item	MS	ZSTD	MS	ZSTD
W6R	1.34	5.6	1.38	6.4
W14	1.28	5.4	1.37	6.2
W2R	1.26	4.4	1.33	4.6
W4R	1.20	2.6	1.26	3.1
W12	0.96	-0.6	0.91	-0.7
W8	0.86	-2.8	0.93	-1.4
W15	0.93	-0.7	0.77	-1.3
W13	0.90	-2.4	0.88	-2.4
W1A	0.84	-2.0	0.60	-3.2
W7	0.83	-4.2	0.81	-4.6
W1	0.79	-2.9	0.72	-2.1
W3	0.77	-3.4	0.65	-3.2

Note. MS represents mean square and ZSTD represents Z-standardized.

For a visual representation of the item difficulty and person ability found for the unidimensional model, refer to Figure 7 in the item-person map. The persons tended towards the lower end of the item-person map indicating a proclivity for positive responses to the items. The items on the map produced some spread to suggest some difference in item difficulty.

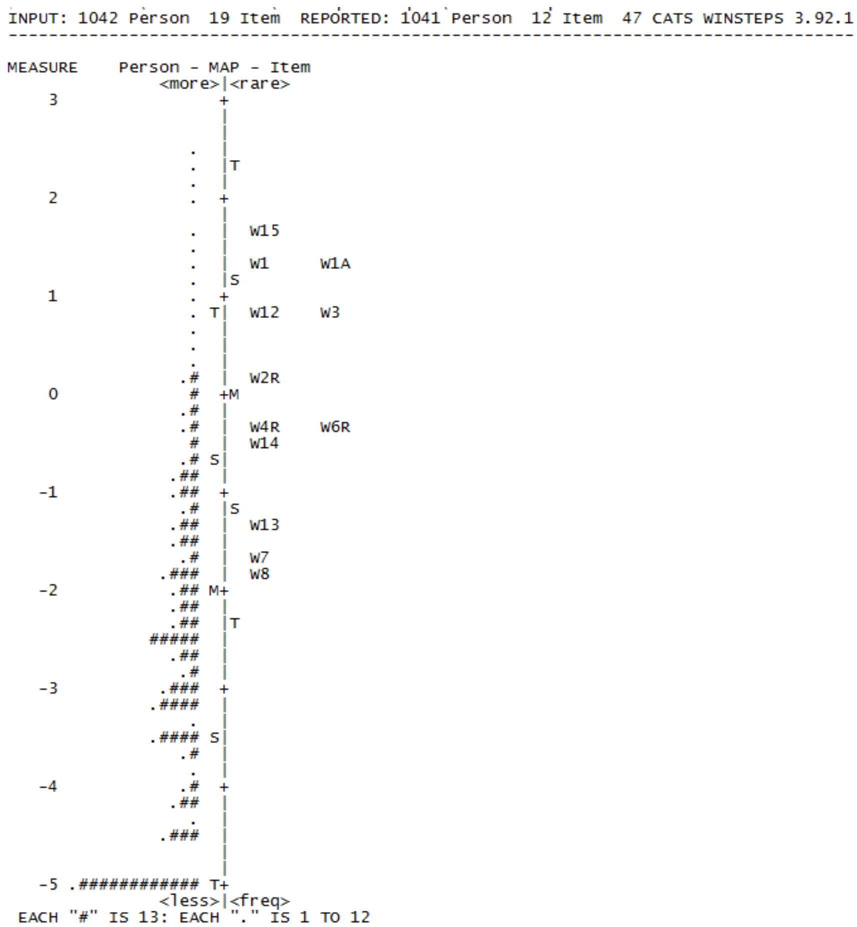


Figure 7. Item-person map for the unidimensional model.

Consecutive approach

Once the unidimensional model was finalized, all 39 items were utilized within the consecutive approach to further explore dimensionality. The same 20 items from the unidimensional analysis were dropped for the consecutive approach due to a lack of data. From this point, 19 items were retained for analysis with the consecutive approach. The consecutive approach began with the results of the initial run of the data for the unidimensional model. Despite the original first contrast eigenvalue being below 2.0, seven of the 19 items misfit. Also, support for a second dimension with eigenvalues < 2.0 can be considered for larger sample sizes (Raiche, 2005). Thus, additional trimming of the unidimensional model, guided by an examination of item fit statistics and the first contrast item indices, established a core dimension. All of the original sample subjects (n=1040) were utilized at the start of the consecutive approach. For this dimension, Dimension A, six items were found to be the best fitting, Table 6. The item infit MS ranged from 0.77 to 1.22 with an overall average infit mean square of 0.99. The mean item outfit MS ranged from 0.76 to 1.24 with an overall average outfit mean square of 0.96.

Table 6
Item Fit Information for Dimension A in Consecutive Approach

	Infit		Outfit	
Item	MS	ZSTD	MS	ZSTD
W14	1.21	4.1	1.24	4.5
W4R	1.22	2.7	1.20	2.6
W6R	1.02	0.4	0.98	-0.3
W1A	0.90	-1.3	0.78	-2.1
W15	0.85	-1.7	0.76	-1.7
W12	0.77	-3.9	0.78	-2.2

Note. MS = mean square and ZSTD = Z-standardized.

Dimension A resulted in a total variance explained by the measure of 56.8%, with 38.7% explained by the person and 18.1% explained by the items. The person separation for this dimension was 0.72 with a reliability of 0.34 and a Cronbach's Alpha of 0.85. While determining the items for Dimension A, six respondents were removed from this analysis, due to person misfit with infit mean squares of over 9.0. The item separation was 11.89 and the model reliability was 0.99. Figure 8 provides the item-person map for Dimension A to display the resulting item difficulty and person ability. The item-person map indicated some spread for both the items and the persons. The placement of the persons on this map revealed a tendency of the participants to respond positively to the items.

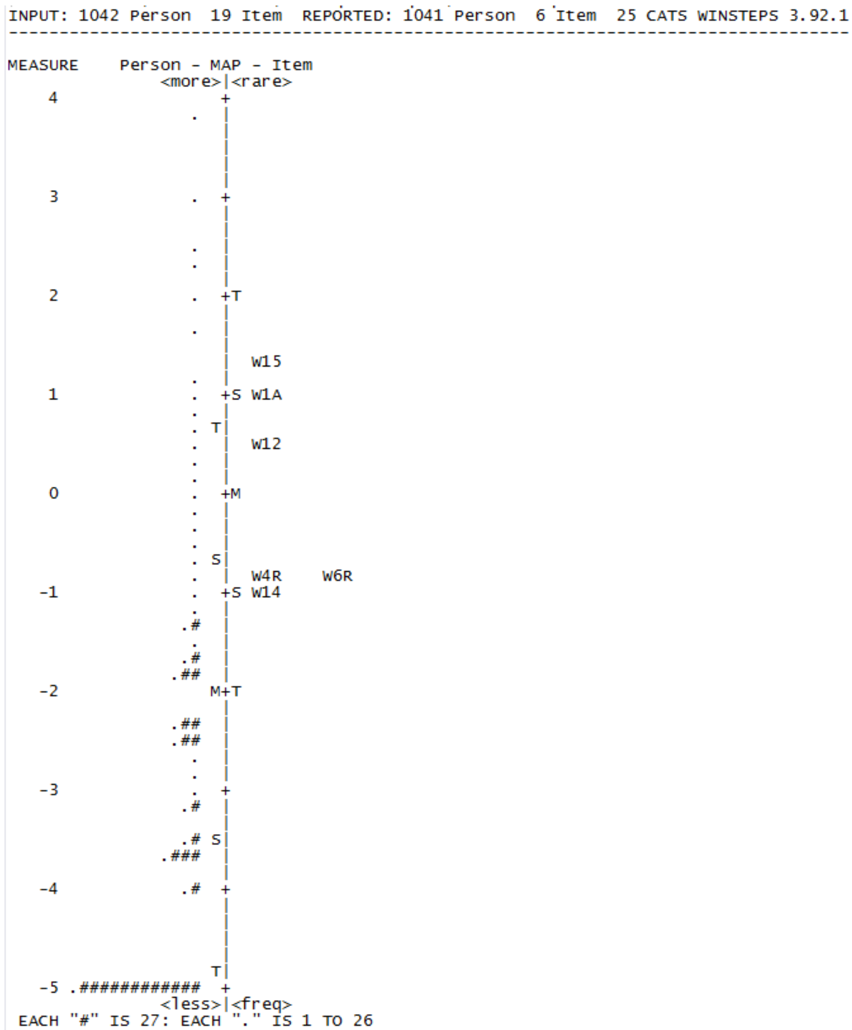


Figure 8. Item-person map for dimension A.

Dimension B was identified by analyzing the remaining thirteen items with Winsteps. The same respondents removed from the Dimension A analysis were removed from the analysis as well. All other respondents were retained while determining the Dimension B items. The thirteen items, beginning with 61.5% of the total variance explained by the measure and the first contrast eigenvalue of 1.79, explained 5.3% of the unexplained variance. As before, the items with a mean square infit of >1.4 were removed. For this dimension, seven items were found to be a poor fit. The item outfit mean square ranged from 0.68 to 1.54, with one exception, and an overall outfit mean square of 0.95 (see Table 7). The mean square outfit maximum for item W1 was just outside of the cutoff criteria, with MS = 0.54, but the item was kept, since further trimming produced less stable models. The item infit MS ranged from 0.75 to 1.33 with an overall infit mean MS of 0.96.

Table 7
Item Fit Information for Dimension B in Consecutive Approach

	Infit		Outfit	
Item	MS	ZSTD	MS	ZSTD
W2R	1.33	5.6	1.54	7.5
W13	1.10	2.1	1.11	1.9
W8	0.94	-1.0	0.98	-0.3
W7	0.88	-2.6	0.85	-2.9
W1	0.78	-3.1	0.54	-3.1
W3	0.75	-3.8	0.68	-2.5

Note. MS = mean square and ZSTD = Z-standardized.

Analysis for Dimension B resulted in six fitting items with a total variance explained of 65.3%, with 39.1% explained by the person and 26.3% explained by the items.

The person separation for this dimension was 1.60 with a reliability of 0.72 and a Cronbach’s Alpha of 0.83. The item separation was 22.36 and the model reliability was 1.00. In addition, no item category recalibration was needed for this dimension. Figure 9 provides the item-person map for Dimension B. Similar to Dimension A, the person placements on the item-person map suggested consistent positive responses. The map revealed that the items of Dimension B were more spread out and may be more difficult to elicit positive responses than those of Dimension A.

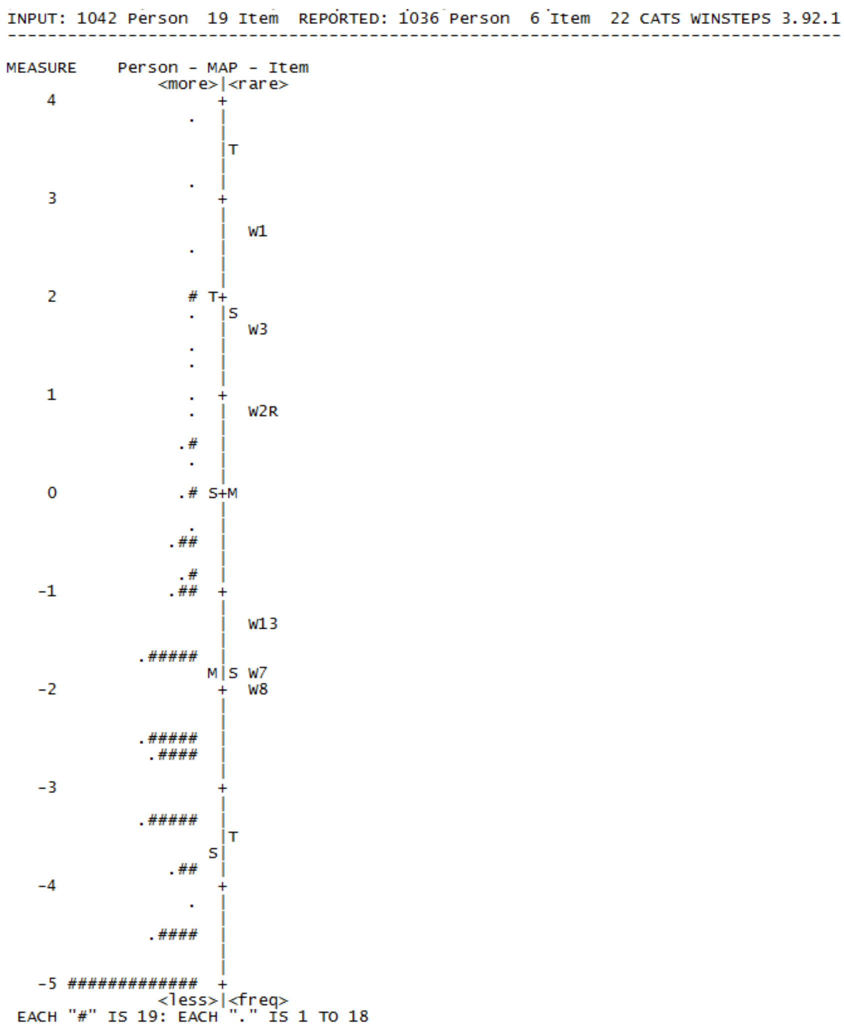


Figure 9. Item-person map for dimension B.

The seven items excluded from Dimensions A and B were the same items excluded from the unidimensional model as well. All of the respondents were retained for the initial analysis of this dimension. Twenty-eight different respondents were removed, while tuning the model. The initial model statistics for these items were total variance explained of 73.4% and the first contrast had an eigenvalue of 1.76, which explained 6.7% of the unexplained variance. As before, the items with a mean square infit of >1.4 were removed. There was no item category calibration needed for this dimension. In this dimension, only one item was found to be poorly fitting. Analysis for Dimension C resulted in six fitting items with a total variance explained of 78.4%, with 42.3% explained by the person and 26.1% explained by the items. The person separation for this dimension was 0.0 with a reliability of 0.0 and a Cronbach's Alpha of 0.54. The item separation was 12.25 and the model reliability was 0.99. See Table 8 for model fit statistics. The item infit MS ranged from 0.83 to 1.38 with an overall mean infit MS of 1.07. The item outfit mean square ranged from 0.59 to 1.90 with an overall mean square outfit of 1.13. As before, the out of range mean square outfit item, W17B, was kept due to model stability. During the final step of the consecutive approach process, item W18 was removed from the model due to both fit indices being outside the cutoff ranges with a MS infit of 2.35 and a MS outfit of 4.68. W18 was not used in any of the models.

Table 8
Item Fit Information for Dimension C in Consecutive Approach

	Infit		Outfit	
Item	MS	ZSTD	MS	ZSTD
W17AR	1.38	1.8	1.90	2.2
W17	1.35	2.8	1.82	2.4
W16	1.03	0.3	0.80	-0.6
W9	0.91	-1.3	0.93	-0.8
W17B	0.90	-0.2	0.59	-0.5
W5	0.83	-3.3	0.77	-4.0

Note. MS represents mean square and ZSTD represents Z-standardized.

The item-person map for Dimension C, as shown in Figure 10, reveals a greater spread of the items and persons than the previous dimensions and the unidimensional model. The person distribution indicates that the respondents tend towards very positive responses. The distribution of the items on this map suggested that the item difficulty was more diverse for this dimension with W16 the most difficult to provide a positive response.

```
INPUT: 1042 Person  19 Item  REPORTED: 1014 Person  6 Item  20 CATS WINSTEPS 3.92.1
--------------------------------------------------------------------------------

MEASURE    Person - MAP - Item
                  <more>|<rare>
    6                .  +T
                        |
                        |
                        |
                        |
    5                .  +
                        |
                     .  |  W16
                    .# T|
                        |
    4                .  +
                        |
                    .#  |
                    .#  |
                        |
    3                .  +S
                   .##  |
                        |
                        |
                        |
    2                  S+
         .############  |
                        |
             .########  |
                        |  W5
    1              .##  +
                        |
                        |
                        |  W9
              ########  |  W17B
    0                   +M
                        |
                       M|
               .######  |
                        |
   -1                .  +  W17AR
                        |
            .#########  |
                     .  |
                        |
   -2           .#####  +
                        |
                   .##  |
                       S|
                        |
   -3                   +S
                    .#  |
                     .  |
                        |
                        |
   -4                   +
              .#######  |
                     .  |
                    .#  |
                    .#  |
   -5                . T+
                        |
                        |
                        |  W17
                        |
   -6                .  +T
                  <less>|<freq>
 EACH "#" IS 13: EACH "." IS 1 TO 12
```

Figure 10. Item-person map for dimension C.

The final model fit and overall item MS statistics for the unidimensional model and each of the identified dimensions from the consecutive approach can be compared using Table 9 and Table 10, respectively.

Table 9

Model Dimensionality and Person/Item Separation and Reliability Comparison

	Unidimensional	Dimension A	Dimension B	Dimension C
Explained Variance				
Total	56.4%	56.8%	65.3%	78.4%
Persons	37.0%	38.7%	39.1%	42.3%
Items	19.3%	18.1%	26.3%	26.1%
First Contrast				
Eigenvalue	1.87	1.46	1.14	1.80
% of Variance	6.8%	10.5%	6.6%	6.5%
Person				
Logit Mean	-2.60	-2.76	-2.75	-0.30
Separation	1.61	0.72	1.60	0.0
Reliability	0.72	0.34	0.72	0.0
Cronbach's Alpha	0.90	0.85	0.83	0.54
Item				
Separation	15.81	11.89	22.36	12.25
Reliability	1.00	0.99	1.00	0.99

Note. Results were determined through the use of Winsteps 3.92.1.

Table 10
Item Mean Square Infit and Outfit Statistics

	Unidimensional	Dimension A	Dimension B	Dimension C
Mean Square Infit				
Minimum	1.34	0.77	0.75	0.83
Maximum	0.77	1.22	1.33	1.38
Mean	1.00	0.99	0.96	1.07
Mean Square Outfit				
Minimum	0.65	0.76	0.68	0.59
Maximum	1.38	1.24	1.54	1.90
Mean	0.97	0.96	0.95	1.13

Note. Infit and outfit score comparison across models.

The resulting analysis established 12 items in the unidimensional model, six items in Dimension A of the consecutive model, six items in the Dimension B of the consecutive model, and six items in the Dimension C of the consecutive model. The item list is shown in Table 11 and the specific items are shown in Appendix A.

Table 11
Items within Dimensions Found in Winsteps

Unidimensional	Dimension A	Dimension B	Dimension C
W1, W1A, W2R, W3, W4R, W6R, W7, W8, W12, W13, W14, W15	W1A, W4R, W6R, W12, W14, W15	W1, W2R, W3, W7, W8, W13	W5, W9, W16, W17, W17AR, & W17B

Note. Dimensions A and B were nested within the unidimensional model.

Upon examining the item text, each of the dimensions were labeled and these labels are listed in Table 12.

Table 12
Consecutive Dimension Labels

Dimension	**Label**
Unidimensional	Parent-Child Relational Well-being
A	Child's Interactions with Others
B	Parent's Expectations
C	Adoption Experience

Note. Labels given to emergent dimensions from the consecutive process.

Multidimensional models

From the consecutive approach (Table 12), three dimensions were determined and used in the multidimensional analysis through ConQuest. The multidimensional approach incorporated the interrelationships of the dimensions as an aspect of model fit that the consecutive approach is incapable of considering. Two items from the third dimension did not fit when utilizing ConQuest. Therefore, those items were removed and the analysis was rerun. Throughout this study, ConQuest took an extended time to converge. The settings used in order to obtain results from ConQuest were relaxed to allow convergence. Because of the low eigenvalue found during the consecutive approach for Dimension C and the loss of two items from this dimension, a 2-dimensional model was also examined in addition to the unidimensional model. An advantage of the 2-dimensional model (Dimension A and Dimension B) was that it was nested within the unidimensional model, which made it easier to compare directly the fit of the models, both models containing 12 items. The items retained for each dimension of the two multidimensional models are listed in Table 13.

Table 13
Items within Dimensions for Multidimensional Models

	Dimension label	No. of items	Items
3-Dimensional Model	Child's Interactions with Others	6	W1A, W4R, W6R, W12, W14, W15
	Parent's Expectations	6	W1, W2R, W3, W7, W8, W13

	Adoption Experience	4	W5, W9, W16, W17
2-Dimensional Model	Child's Interactions with Others	6	W1A, W4R, W6R, W12, W14, W15
	Parent's Expectations	6	W1, W2R, W3, W7, W8, W13

Note. List of items in each of the dimensions for the examined models.

In the tested 3-dimensional model, the weighted mean squares ranged from 0.85 to 1.04. The separation reliability was .95. In Table 14, a high correlation was found between Dimension A and Dimension B with *r* =.80. There was almost no correlation found between Dimension C and the other dimensions.

Table 14
Multidimensional Model 3-Dimensions Covariance – Correlation Matrix

Dimension	Dim A	Dim B	Dim C
Child's Interactions with Others (Dim A)		0.33	0.09
Parent's Expectations (Dim B)	.80		0.19
Adoption Experience (Dim C)	.17	.28	

Note. Covariance above the diagonal and correlation below the diagonal space.

The item-person map for the 3-dimensional MIRT model revealed a slight disconnect between the item difficulty and person ability to respond positively to items, in Figure 11 for all of the dimensions. A large spread was delineated for items when examining all three dimensions.

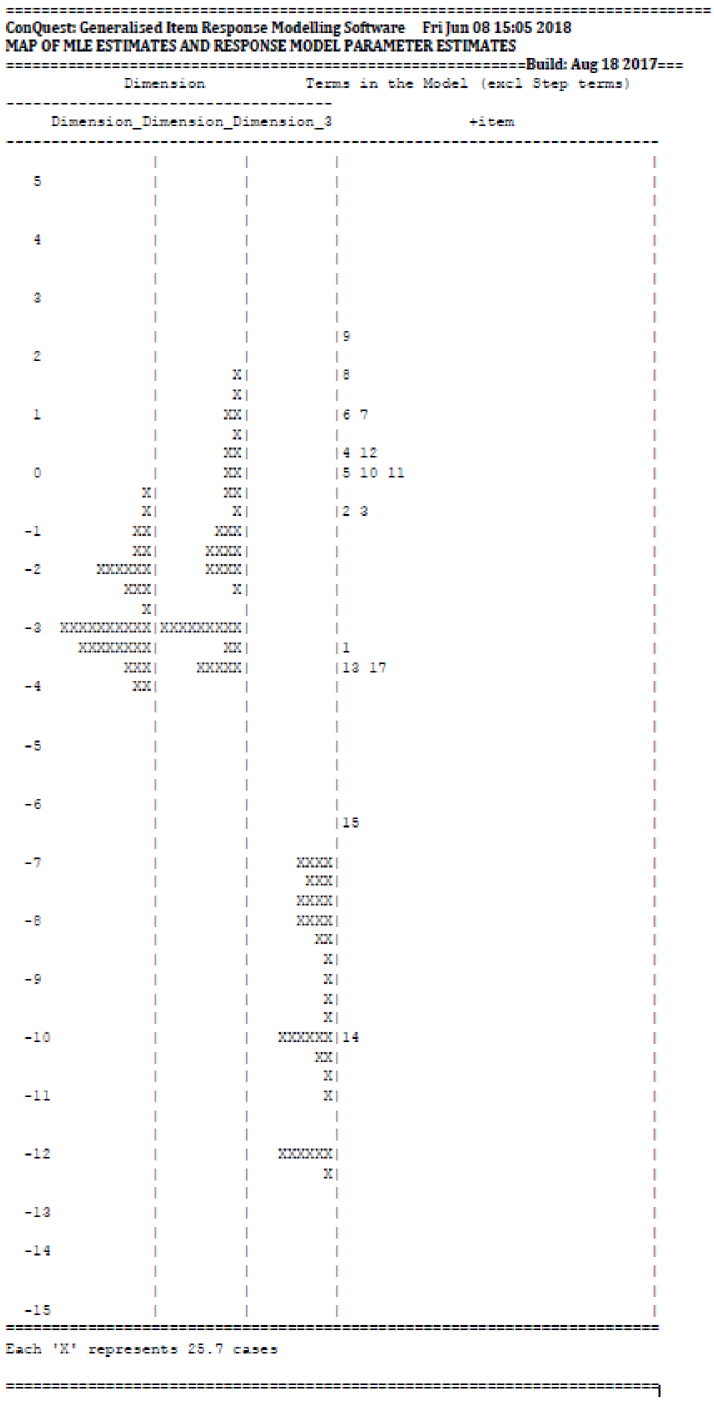


Figure 11. Item-person map for 3-dimensional MIRT model. Dimension 1 bolded and dimension 2 italicized, and dimension 3 in an unadjusted type set.

In the 2-dimensional MIRT model, the weighted mean squares ranged from 0.85 to 1.18. The separation reliability was .975. In Table 15, a low correlation between Dimension A and Dimension B with *r* =.33 was observed.

Table 15
Multidimensional Model (2-dimensions) Covariance – Correlation Matrix

Dimension	Dim A	Dim B
Child's Interactions with Others (Dim A)		.28
Parent's Expectations (Dim B)	0.33	

Note. Covariance above the diagonal and correlation below the diagonal.

Inspection of the item-person map for the 2-dimensional model showed clustering of persons at the lower regions of the map, indicating it was less difficult to give a positive response to items in both dimensions, see Figure 12. The positioning of the persons and items affirmed the idea that the persons agreed with the items.

```
==============================================================================
ConQuest: Generalised Item Response Modelling Software    Fri Feb 15 11:32 2019
MAP OF MLE ESTIMATES AND RESPONSE MODEL PARAMETER ESTIMATES
==========================================================Build: Aug 18 2017===
        Dimension      Terms in the Model (excl Step terms)
--------------------------
     Dimension_Dimension_2              +item
--------------------------------------------------------------
   6            |         |                                  |
                |         |                                  |
                |         |                                  |
   5            |         |                                  |
                |         |                                  |
                |         |                                  |
   4            |         |                                  |
                |         |                                  |
                |         |                                  |
   3            |         |                                  |
                |         |                                  |
                |         |                                  |
   2            |         |                                  |
                |         |                                  |
                |         |9                                 |
   1            |        X|6                                 |
                |         |7 8                               |
                |        X|                                  |
                |         |1 4                               |
   0            |       XX|                                  |
               X|        X|5                                 |
               X|        X|2 12                              |
  -1          XX|       XX|3 10                              |
               X|         |11                                |
               X|        X|                                  |
  -2         XXX|        X|                                  |
          XXXXXX|        X|                                  |
               X| XXXXXXXX|                                  |
  -3        XXXXXXXXXXXXXX|                                  |
      XXXXXXXXXX|    XXXXX|                                  |
             XXX|         |                                  |
  -4           X|         |                                  |
                |         |                                  |
                |         |                                  |
  -5            |         |                                  |
==============================================================
Each 'X' represents 29.4 cases

==============================================================
```

Figure 12. Item-person map for 2-dimensional MIRT model. Dimension 1 bolded and dimension 2 in an unadjusted typeface.

When the overall fit of the models was examined, it was determined that the combination of Consecutive Dimensions A and B models were the best fitting model (see Table 16). The 3-dimensional model had an AIC of 23024.11, the combined dimensions of the consecutive approach had a combined AIC of 21460.14, and the unidimensional model had an AIC of 17866.46 which indicated the 2-dimensional model showed the best fit with an AIC of 17227.00. Deviance (G^2) was also lowest for the 2-dimensional model. Only models comprising the same item set can be directly compared. The 2-dimensional model out performed both the unidimensional model and the 3-dimensional model with lower G^2 and AIC. When the final comparison was made between the 2-dimensional MIRT model and the combination of Consecutive Dimensions A and B models through the use of AIC, Akaike (1985) suggests AIC can be compared for nonnested models, combined models of Dimension A and Dimension B were selected due to the lowest AIC of 16745.93.

Table 16
Comparison of Model Fit Between All Models

	Sample Size	Parameters Estimated	G^2	AIC
Unidimensional	1040	37	17792.45	17866.46
Consecutive – Dimension A	1040	20	7783.23	7823.23
Consecutive – Dimension B	1040	18	8886.70	8922.70
Consecutive – Dimension C	1040	16	4682.21	4714.21
Consecutive – Combined	1040	38	16669.93	AIC_net =

Dimensions A and B				16745.93
Multidimensional 3-Dims	1040	53	22619.17	22725.17
Multidimensional 2-Dims	1040	39	17149.00	17227.00

Note. The combined Dimensions A & B from the consecutive approach were determined to be the best fitting model.

Research Question Two

Were scale response categories used appropriately for each of the included items?

The first item found to be misaligned using Andrich Thresholds was one that asked how the child felt about being adopted (W14). It was determined that the categories needed to be collapsed to decrease noise found within categories 3, 4, and 5. Figure 13 reveals the initial overlap between the categories 3, 4, and 5, while Figure 14 shows the category probability plots once the categories were merged.

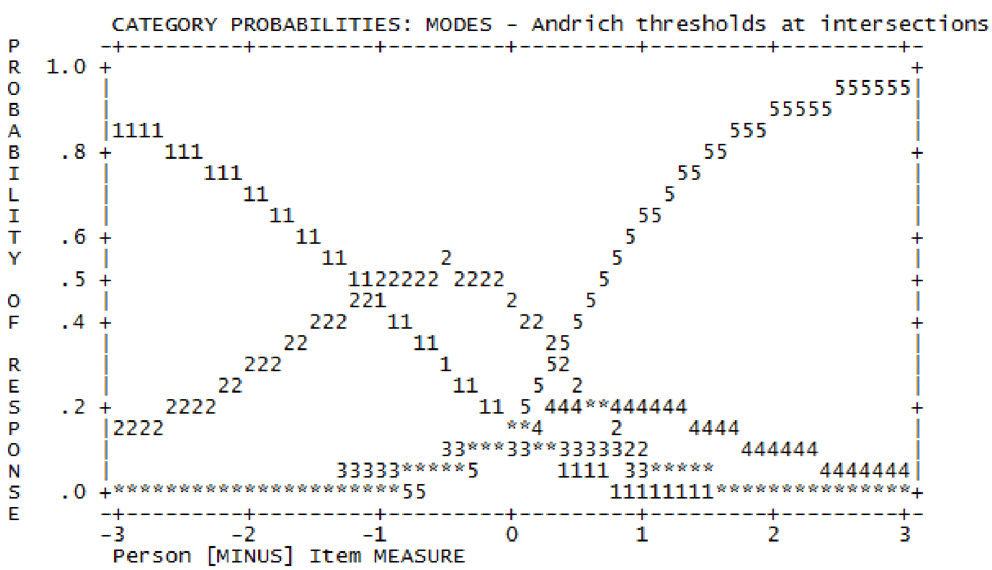


Figure 13. Category map and Andrich Thresholds for W14 pre-adjustments.

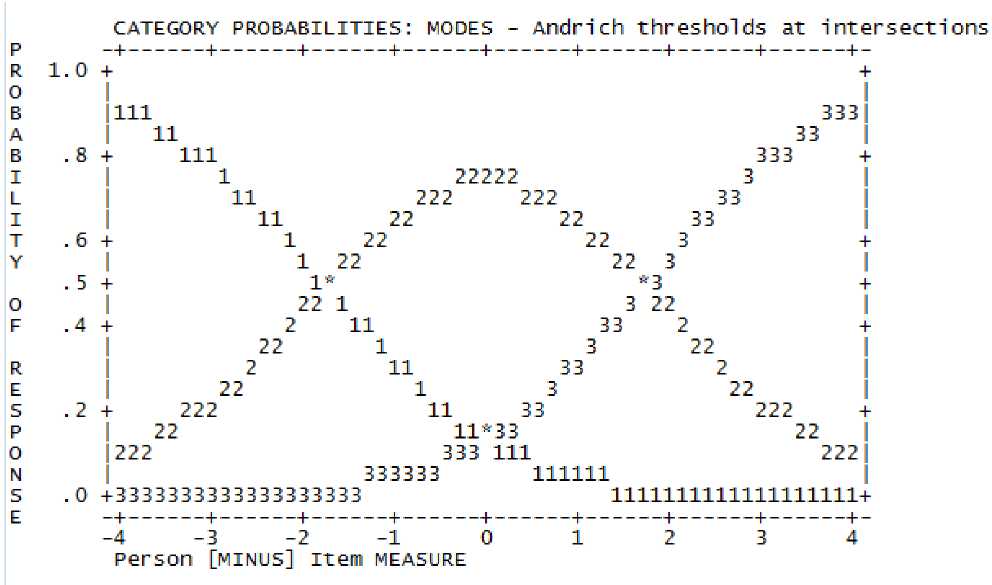


Figure 14. Category map and Andrich Thresholds for W14 post-adjustments.

After the adjustments to item W14, the fit statistics were improved to an acceptable level and the item was retained.

Item W4R, asking parent if the parent and child make life decisions together, was the second item found having overlapping categories between the fourth and fifth categories, Figure 15.

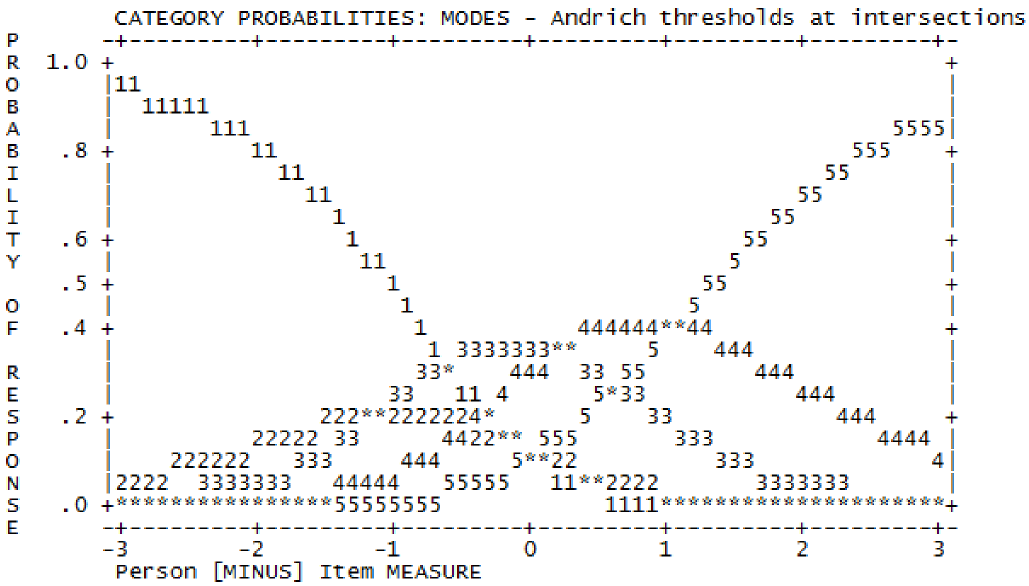


Figure 15. Category map and Andrich Thresholds for W4R pre-adjustments.

The fourth and fifth categories were merged (see Figure 16) which allowed this item to remain due to improved item fit statistics. No other items showed problems with respondent use of the response scale.

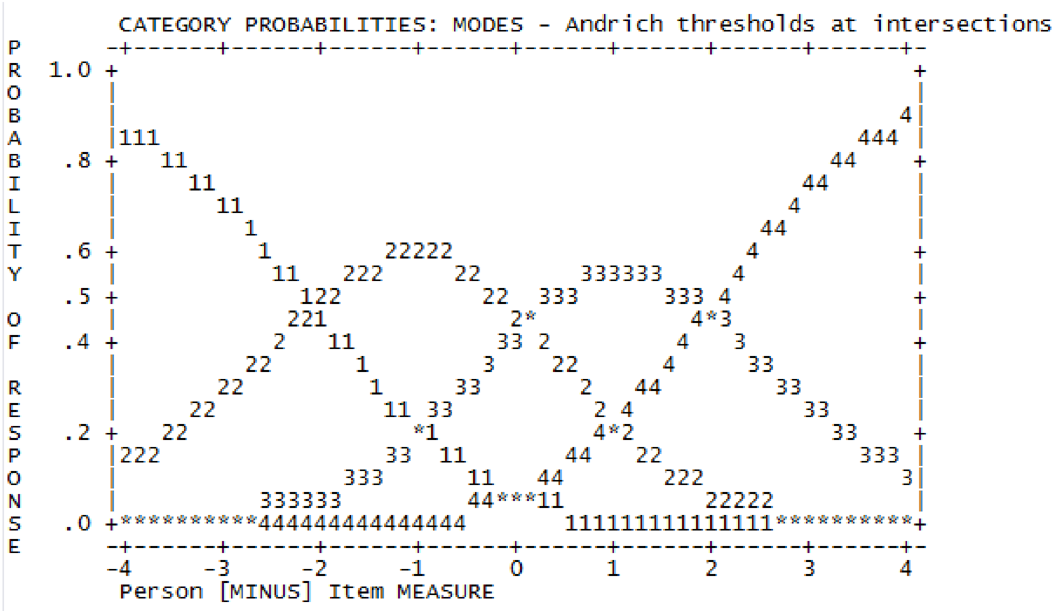


Figure 16. Category map and Andrich Thresholds for W4R post-adjustments.

Research Question Three

Did respondents for different adoption types (inter-country, domestic-private, and foster-to-adopt) interpret the items differently as observed through differential item functioning (DIF)? How did the differential item functioning results for adoption type compare between the unidimensional, the consecutive, and the multidimensional approach?

Each item within the models was compared across the adoption type groups of inter-country, foster-to-adopt, and domestic-private. DIF was used to determine if there was invariance between the groups for each item. The assumption that each of the adoption groups responded similarly across the items held for the majority of the items. Invariance was identified when the logit position difference was ≤ .50 with a $p \leq .01$.

Items W2R, W12, W14, and W15 met the criteria that revealed a response difference between the adoption types within the unidimensional approach and the consecutive approach (see Table 17). Item W2R was easier for respondents from the inter-country adoption type to provide a positive response to than for the foster-to-adopt groups for both the unidimensional model and Dimension B of the consecutive approach. The DIF analysis indicated that inter-country adoption participants found it easier to respond positively to item W12 than domestic-private participant for both the unidimensional model and Dimension A of the consecutive approach, while the research showed that item W12 was easier only for the inter-country adoption group than the foster-to-adopt group when examining Dimension A and not the unidimensional model. Foster-to-adopt participants answered more positively to item W14 than both the inter-country and domestic-private groups for both the unidimensional model and Dimension A. Finally, positive responses for item W15 were easier for the domestic-private respondents in both the unidimensional model and Dimension A than for the foster-to-adopt participants, according to the DIF analysis. DIF results were found to be very similar, whether the items were combined into one dimension (unidimensional) or separated via the consecutive approach. Only item, W15 was found to be a possible concern between the responses of the foster-to-adopt and domestic-private parents, due to the distance between the estimates, for the MIRT models, however, no significant difference was found through the multidimensional DIF analysis.

Table 17
Differential Item Functioning for unidimensional and Consecutive Approaches

Item	Model/Dimension	Lower Logit Position	Higher Logit Position	*p*
W2R	Unidimensional	Inter-country	Foster-to-adopt	<.001
W2R	Dimension B	Inter-country	Foster-to-adopt	<.001
W12	Unidimensional	Inter-country	Domestic-private	<.001
W12	Dimension A	Inter-country	Domestic-private	<.01
W12	Dimension A	Inter-country	Foster-to-adopt	<.001
W14	Unidimensional	Foster-to-adopt	Inter-country	<.01
W14	Unidimensional	Foster-to-adopt	Domestic-private	<.001
W14	Dimension A	Foster-to-adopt	Inter-country	<.001
W14	Dimension A	Foster-to-adopt	Domestic-private	<.001
W15	Unidimensional	Domestic-private	Foster-to-adopt	<.01
W15	Dimension A	Domestic-private	Foster-to-adopt	<.01

Note. DIF for these approaches examined, DIF contrast >.50.

Research Question Four

How did the person logits and item difficulty compare for each dimension found within models from the unidimensional, consecutive and multidimensional approaches compare across models?

a. Did the person logit positions correlate across dimensions/models and software packages?

b. Using a cluster of three independent variables, Adoptive Family with or without Biological Children, Child Lived with Birth Family, and Adoptive Parent/s and Child of Differing Races, as predictors in a regression analysis, were the R^2 values comparable between the models and the software?

c. Using the same cluster of family characteristic variables as independent variable in a canonical correlation analysis, were the canonical R^2 values comparable between the models and the software?

d. How did the item difficulty compare across the models examined?

In order to examine differences across the software packages Winsteps and ConQuest, the person logit positions were correlated (see Table 18). The unidimensional model person logit positions for Winsteps and Conquest were correlated positively at $r = .76$, $p \leq .01$, a strong association. Both dimensions of the 2-dimensional model correlated strongly with both unidimensional models, producing correlations of $r = .83$ and $r = .79$, $p \leq .01$, for the Winsteps unidimensional model; and $r = .95$, $p \leq .01$ and $r = .87$, $p \leq .01$, for the ConQuest unidimensional model. The dimensions within the 3-dimensional model did not correlate well with the other models, except within the third dimension. However, the third dimension of the 3-dimensional model contained different items than the other models that resulted in moderate correlations. Both dimensions within the 2-dimensional model generated strong correlations with both Winsteps and ConQuest unidimensional models. The 2-dimensional model maintained strong correlations with the consecutive dimensions A & B, as well (see Appendix E). Dimension C had a strong correlation between the Winsteps and Conquest models, $r = .92$, $p \leq .01$, but produced weak correlations ranging from $r = .32$ to $r = .36$, $p \leq .01$, with the majority of the other dimensions.

Regression analysis was used to compare the consistency between software and each model, with a single outcome, using family characteristics as predictor variables: did the adoptive parent/s have biological child/ren, had the adoptive child lived with their birth family, and was there a racial difference between the adoptive parent and the child. The unidimensional models and the single dimension models were used in this analysis. When the family characteristics were used the R^2 for each of the models were small, ranging from $R^2 < .01$ through .028, as seen in Table 18. All of the testing resulted in significant findings at $p \leq .01$. These results, as a whole, indicated that the cluster of these family characteristics were significant predictors for these models; however, the impact on the models themselves were variable and small. The unidimensional model explained the most variance when using this comparison with the Family Characteristics cluster of independent variables, due to the largest R^2 identified, $R^2 = .028$.

Table 18
R^2 Comparisons of Observed Models to Family Characteristics

Model	R^2	Std. Error of Est.	F	Sig.
Unidimensional. – Winsteps	.028	1.82	10.03	<.001
Unidimensional. – ConQuest	.019	0.22	6.13	<.001
Dim. A - Winsteps	.012	1.60	4.28	.005

Dim. A – ConQuest	.018	0.22	6.26	<.001
Dim. B - Winsteps	.024	2.32	8.60	<.001
Dim. B - ConQuest	.015	1.59	5.14	.002
Dim. C - Winsteps	.010	2.45	3.52	.015
Dim. C - ConQuest	.012	0.27	4.06	.007

Note. Multiple regression analysis used with dimensional person logit positions as the dependent variable and the Family Characteristics cluster as the three independent variables.

The R^2 differences were used to compare the models. The Winsteps Unidimensional model and The Winsteps Dimension B model accounted for slightly more variance than their ConQuest counterparts (see Table 19). The difference for both comparisons, the unidimensional and Dimension B, had a R^2 difference of .009, which was the largest difference when comparing similar models. The ConQuest models for Dimension A and Dimension C were seen to have a greater difference than the Winsteps versions of these similar models/dimensions.

Table 19
Difference in R^2 between Models Winsteps by ConQuest by Family Characteristics

	ConQuest			
Model	Unidimensional	Dim A	Dim B	Dim C
Unidimensional	.009			
Dimension A		-.006		
Dimension B			.009	

Dimension C	-.002

Note. Left column represents Winsteps and top row represents ConQuest models. Differences determined by subtracting the top row model from the column model. Negative findings indicated that top row model stronger than column model.

In order to analyze the multidimensional models by the Family Characteristics cluster, comparisons were done using a series of canonical correlations. From these analyses, the canonical R^2 was used as the indicator. The multidimensional models, as well as, combinations of the dimensions established through the consecutive approach were evaluated (see Table 20). Through this analysis the 3-dimensional model produced the largest canonical R^2, explaining the greatest amount of variability.

Table 20
Canonical R^2 Comparisons of Observed Models to Family Characteristics

Model	Canonical R^2	F	Sig.
Combined Winsteps Dimension A, B, & C	.024	3.81	<.001
Combined ConQuest Dimension A, B, & C	.022	3.39	<.001
3-Dimensional – ConQuest	.025	3.15	.001
Combined Winsteps Dimension A & B	.024	4.56	<.001
Combined ConQuest Dimension A & B	.019	3.91	.001
2-Dimensional – ConQuest	.023	4.15	<.001

Note. Canonical correlation analysis used for the multivariate dependent variables and the Family Characteristics cluster as the three variable independent variable.

Within the 3-dimensional comparisons, the 3-dimensional model was larger by .001 than the Winsteps combined consecutive dimensions ABC and larger by .003 than the ConQuest combined consecutive dimensions ABC, in Table 21. For the 2-dimensional models, the Winsteps combined consecutive dimensions AB yielded the larger R^2 with positive differences of .001 and .005 to the 2-dimensional model and the ConQuest combined consecutive dimensions AB, respectively.

Table 21
Difference in Canonical R^2 between Models Winsteps by ConQuest by Family Characteristics

Model	Winsteps ABC	ConQuest ABC	3-Dim	Winsteps AB	ConQuest AB	2-Dim
Dim. ABC - W	-	.002	-.001	0.0	.005	.001
Dim. ABC - CQ	-.002	-	-.003	-.002	.003	-.001
3-Dimensional	.001	.003	-	.001	.006	.002
Dim. AB – W	0.0	.002	-.001	-	.005	.001
Dim. AB - CQ	-.005	-.003	-.006	-.005	-	-.004
2-Dimensional	-.001	.001	-.002	-.001	.004	-

Note. Differences determined by subtracting the top row model from the column model. Negative findings indicated that top row model stronger than column model. W represents Winsteps and CQ represents ConQuest.

The item difficulties followed a similar pattern for Dimension A, items W1A through W14, for all of the models as seen in Figure 17. The ConQuest models diverged from the Winsteps item difficulty positions in Dimension B, items W7, W8, and W13. The ConQuest models identified these items as more difficult than the Winsteps models.

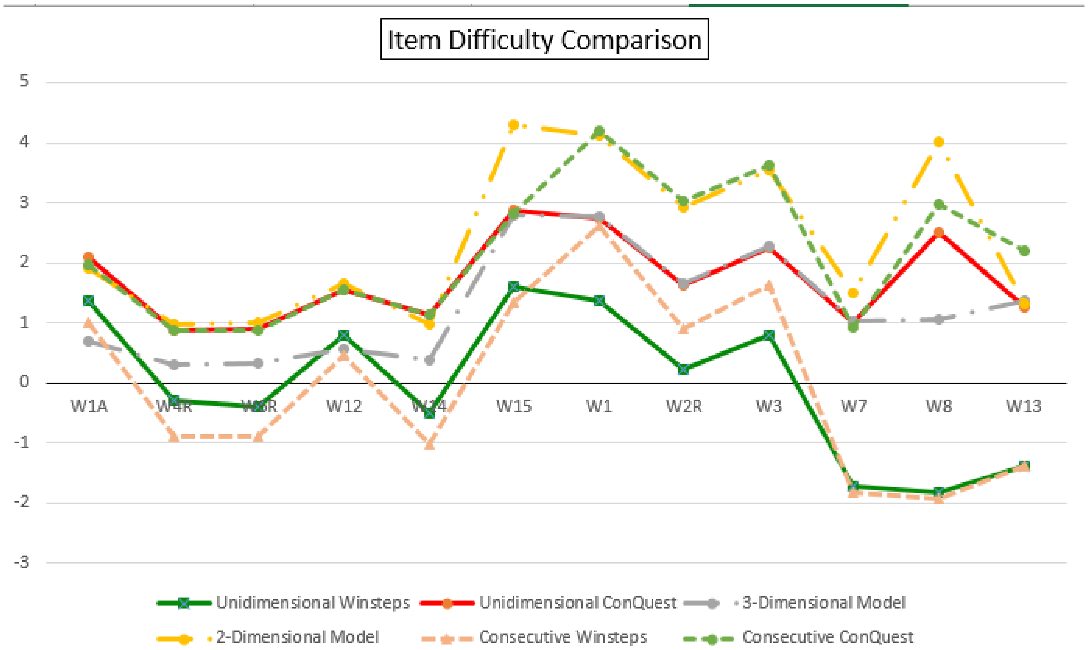


Figure 17. Item difficulty comparisons across the models. Consecutive models combined for comparison.

Through the ConQuest analyses, other information was provided to better understand the quality of the models examined. From the skewness and kurtosis indices, the distribution of person logit position for all of the models was non-normal (Table 22). All but one of the variables showed skewness > 1.0. Dimension C was approximately normal with skewness = 0.65. Similarly, the kurtosis for the models indicated that the data were peaked, particularly the Unidimensional model and two dimensions of the 3-dimensional model with kurtosis scores over 5.0. Dimension C had a kurtosis of -0.57

indicating a flat distribution for this dimension. There was a great deal of data missing from the dataset due to the structured skip protocols, which ranged from 32.6% to 76.4% of the data missing by dimension.

Table 22
Dataset 1 Characteristics

Model	Skewness	Kurtosis	Missing Data
Unidimensional	2.57	6.20	66.9%
Consecutive Dimension A	1.86	3.69	58.8%
Consecutive Dimension B	2.02	3.79	32.6%
Consecutive Dimension C	0.65	-0.57	69.6%
3-Dimensional – A	-0.44	6.40	76.4%*
3-Dimensional – B	1.11	7.41	
3-Dimensional – C	0.87	3.93	
2-Dimensional – A	1.27	2.31	66.9%*
2-Dimensional – B	1.09	2.02	

Note. Majority of these results provided through ConQuest, except for the dimensional scores for the 3-dimension and 2-dimensional models, for these models SPSS was used. * This percentage represents a ConQuest calculation for the whole model.

Research Question Five

With the best fitting model selected, were the dimensionality and model fit replicated in a second half of the dataset for cross validation? Were the item fit, DIF, and validation measures comparable across the two halves of the dataset?

Upon completion of the analysis of the first half, it was determined that the combination of the two consecutive dimensions A and B were the best fitting for the first half of the dataset. It was surmised that these results were due to the large number of missing responses, due to the data collection protocols. The second dataset was prepared for analysis, which included the adjusting of the response scale to two items: how the child felt about being adopted W14, and are life decisions made together, W4R, to match the previous analysis. Figures 18 and 19 display the scale use with adjusted item response scales.

```
FOR GROUPING "0" Item NUMBER: 5  W14

        CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+------+------+------+------+------+------+------+------+-
R  1.0 +                                                         +
O      |                                                         |
B      |11                                                     33|
A      | 111                                               333   |
B   .8 +    11                                           33      +
I      |      11                2222222                33        |
L      |        11           222       222           33          |
I      |          1        22             22        3            |
T   .6 +           11    22                 22    33             +
Y      |             1  2                     2  3               |
    .5 +              *2                       2*                +
O      |             2 11                     33 2               |
F   .4 +           22    1                   3    22             +
       |          2       11               33       2            |
R      |        22          1             3          22          |
E      |      22             11         33             22        |
S   .2 +    22                 11     33                 22      +
P      | 222                     11 33                     222   |
O      |22                      333*111                        22|
N      |                  333333       111111                    |
S   .0 +333333333333333333                   11111111111111111111+
E      -+------+------+------+------+------+------+------+------+-
       -4     -3     -2     -1      0      1      2      3      4
        Person [MINUS] Item MEASURE
```

Figure 18. Category map and Andrich Thresholds for W14 post-adjustment for dataset 2.

```
FOR GROUPING "0" Item NUMBER: 2  W4R
```

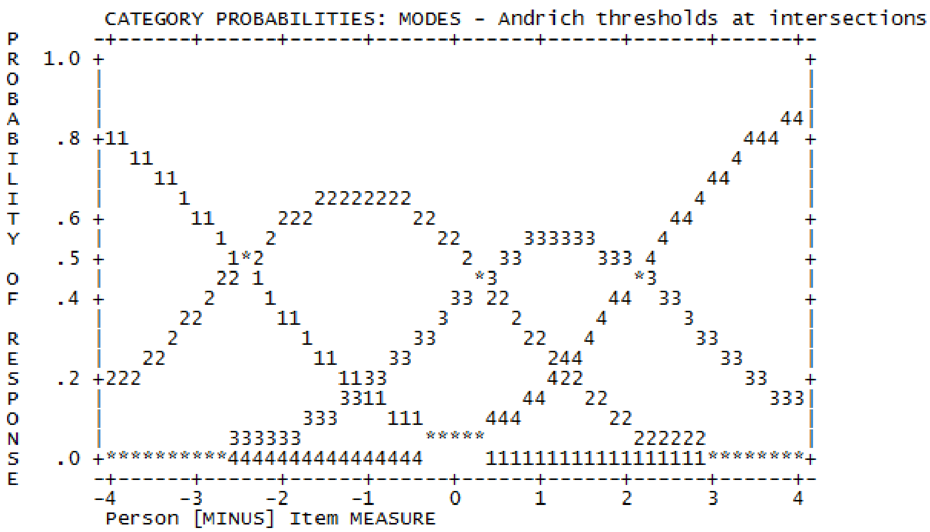


Figure 19. Category map and Andrich Thresholds for W4R post-adjustment for dataset 2.

For Dimension A, the unweighted and weighted mean squares ranged from 0.90 to 1.0 (Table 23), which was within the acceptable bounds of the infit indices. The separation reliability was .97 and test reliability of .176 for Dimension A for dataset 2. These reliability estimates matched closely the findings within ConQuest for Dimension A, separation reliability = .97 and test reliability = .147.

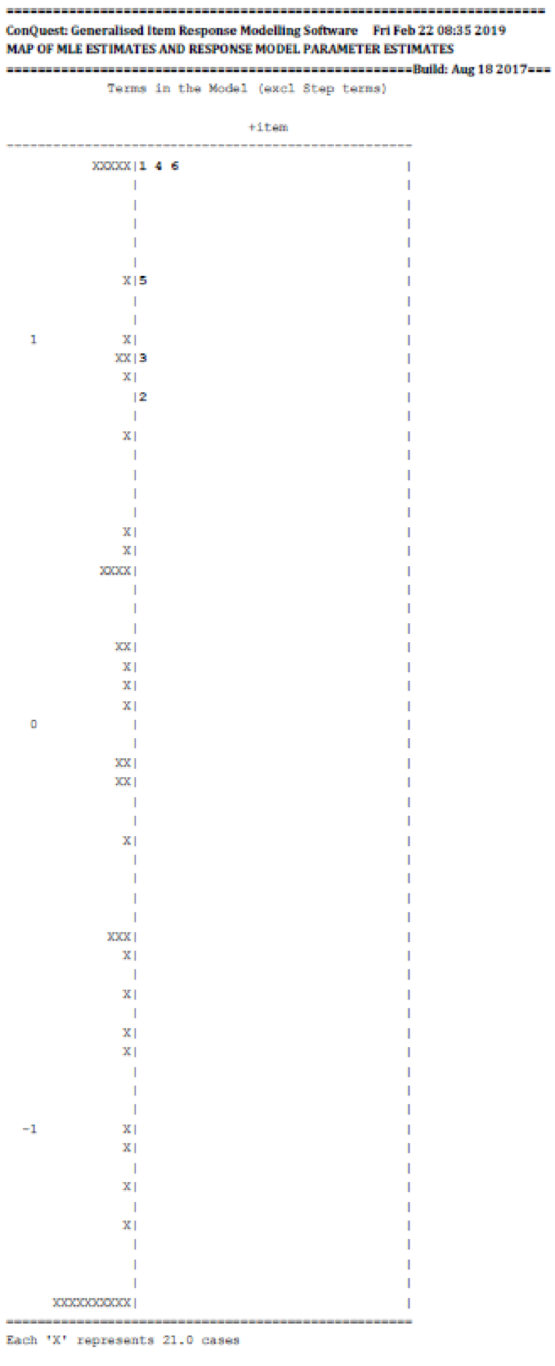
Table 23
Unweighted and Weighted Fit for Dimension A

	Unweighted Mean Square	Weighted Mean Square
W1A	0.90	0.95
W4R	0.91	0.92
W6R	0.89	0.90

W12	0.88	0.94
W14	0.96	0.97
W15	0.91	1.0

Note. Unweighted mean square represented infit and weighted mean square represented outfit.

The item-person map for the verification of Dimension A showed a placement of the items near the top, while the persons were spread throughout the range (Figure 20). The map revealed that the majority of the persons' abilities loading below the item difficulty, similar to the item-person map of the first dataset for Dimension A, Figure 8.



```
==========================================================================
ConQuest: Generalised Item Response Modelling Software    Fri Feb 22 08:35 2019
MAP OF MLE ESTIMATES AND RESPONSE MODEL PARAMETER ESTIMATES
=======================================================Build: Aug 18 2017===
              Terms in the Model (excl Step terms)

                                    +item
---------------------------------------------------------
           XXXXX|1 4 6                                  |
                |                                       |
                |                                       |
                |                                       |
                |                                       |
                |                                       |
               X|5                                      |
                |                                       |
                |                                       |
   1           X|                                       |
              XX|3                                      |
               X|                                       |
                |2                                      |
                |                                       |
               X|                                       |
                |                                       |
                |                                       |
                |                                       |
                |                                       |
               X|                                       |
               X|                                       |
            XXXX|                                       |
                |                                       |
                |                                       |
                |                                       |
              XX|                                       |
               X|                                       |
               X|                                       |
               X|                                       |
   0            |                                       |
                |                                       |
              XX|                                       |
              XX|                                       |
                |                                       |
                |                                       |
               X|                                       |
                |                                       |
                |                                       |
                |                                       |
                |                                       |
             XXX|                                       |
               X|                                       |
                |                                       |
               X|                                       |
                |                                       |
               X|                                       |
               X|                                       |
                |                                       |
                |                                       |
                |                                       |
  -1           X|                                       |
               X|                                       |
                |                                       |
               X|                                       |
                |                                       |
               X|                                       |
                |                                       |
                |                                       |
                |                                       |
      XXXXXXXXXX|                                       |
=========================================================
Each 'X' represents 21.0 cases
```

Figure 20. Item map for dimension A dataset 2.

Through ConQuest, the examination of Dimension B returned unweighted and weighted mean squares ranging from 0.90 to 1.08 (Table 24). The separation reliability was .98 and test reliability .54 for Dimension B. Again, these reliability estimates from Dimension B matched closely with the first dataset's consecutive approach findings within ConQuest for Dimension B, separation reliability = .98 and test reliability = .58.

Table 24
Unweighted and Weighted Fit for Dimension B

	Unweighted Mean Square	Weighted Mean Square
W1	0.77	0.90
W2R	1.02	1.08
W3	0.99	0.94
W7	0.97	1.0
W8	0.91	0.96
W13	1.05	0.99

Note. Unweighted mean square represented infit and weighted mean square represented outfit.

A larger range was produced by Dimension B for the item-person map than that produced by Dimension A of dataset 2. The items in Dimension B were found to be on the higher end of the range, while the persons' placements are spread across the range, see Figure 21. As before, the item-person map for this dimension is similar to the original Dimension B item-person map found at Figure 9.

```
===================================================================
ConQuest: Generalised Item Response Modelling Software    Fri Feb 22 08:46 2019
MAP OF MLE ESTIMATES AND RESPONSE MODEL PARAMETER ESTIMATES
==================================================Build: Aug 18 2017===
             Terms in the Model (excl Step terms)

                               +item
----------------------------------------------------
                |                                  |
   5            |                                  |
               X|1                                 |
                |                                  |
               X|                                  |
   4          XX|3                                 |
              XX|                                  |
               X|                                  |
   3          XX|                                  |
               X|2                                 |
               X|6                                 |
      XXXXXXXXXX|                                  |
   2           X|                                  |
                |                                  |
      XXXXXXXXXX|                                  |
             XXX|4 5                               |
   1      XXXXXX|                                  |
            XXXX|                                  |
               X|                                  |
   0         XXX|                                  |
                |                                  |
           XXXXX|                                  |
          XXXXXX|                                  |
  -1          XX|                                  |
              XX|                                  |
              XX|                                  |
            XXXX|                                  |
  -2           X|                                  |
            XXXX|                                  |
                |                                  |
                |                                  |
  -3            |                                  |
====================================================
Each 'X' represents 13.4 cases

====================================================
```

Figure 21. Item map for dimension B dataset 2.

A comparison of the model fit using deviance and AIC revealed that Dimension A of dataset 2 had a lower value than for the first half of the data, with G^2 = 7404.37 and

AIC = 7442.37. Dimension B of the second half of the dataset was higher than Dimension B of the original results, with G^2 = 9095.60 and AIC =9129.60. The combined results of the consecutive approach Dimensions A and B for the second half of the data yielded the lowest results of all of the first half models, G^2 =16499.97 and AIC = 16751.97 (see Table 25).

Table 25
Comparison of Model Fit Between Models for Dataset 1 and Dataset 2

	G^2	AIC	Dataset 1 A+B G^2	Dataset2 A+B AIC
Dataset 1 – Dimension A	7783.23	7823.23	16669.93	16745.93
Dataset 2 – Dimension A	7404.37	7442.37	16499.97	16751.97
Dataset 1 – Dimension B	8886.70	8922.70		
Dataset 2 – Dimension B	9095.60	9129.60		
Unidimensional	17792.45	17866.46		
Multidimensional 2-Dimensions	17149.00	17227.00		

Note. A comparison of model fit across data halves and model types.

Items within the Dimensions A and B were compared across the adoption type, inter-country, foster-to-adopt, and domestic-private, for the second dataset. The same assumptions used for the first half of the NSAP data were used to determine invariance, (logit position difference ≥ .50 with $p \leq .01$). Items W4R and W14 met the criteria that suggested response differences between the adoption types, see Table 26. Item W4R was easier for respondents from the foster-to-adopt adoption type to provide a positive response to than for the domestic-private participants. In the first dataset, item W4R was not found to have differing responses across adoption type. The DIF analysis also indicated that foster-to-adopt adoption participants found it easier to respond positively to item W14 than domestic-private and inter-country participants. The invariance found in the second dataset was consistent with the first dataset for this item. No items in Dimension B were found to have DIF, indicating that the responses to Dimension B were invariant across adoption type for the second half of the data. Item W2R showed DIF in the first half of the data but not in the second half.

Table 26
Differential Item Functioning for Dataset 2

Item	Model/Dimension	Lower Logit Position	Higher Logit Position	*p*
W4R	Dimension A	Foster-to-adopt	Domestic-private	<.001
W14	Dimension A	Foster-to-adopt	Inter-country	<.001
W14	Dimension A	Foster-to-adopt	Domestic-private	<.001

Note. DIF contrast >.50 and $p \leq .01$.

Following the same procedures for the verification of the second half of the data, the person logit scores of Dimension A and Dimension B correlated significantly, $r = .64$, $p = 0.01$. When comparing the dimensions through a multiple regression analysis with the same family characteristics as before, the R^2 results were consistent with previous analyses. Dimension A R^2 for the second half was .03 higher than the ConQuest

Dimension A R^2 result (Table 27). For Dimension B of the second dataset, the R^2 was equal to the Winsteps R^2 Dimension B.

Table 27
Comparison of Dataset 1 and 2 by Family Characteristics

	R^2	Std. Error of Est.	F	Significance
Dimension A Winsteps – Dataset 1	.012	1.60	4.28	.005
Dimension A ConQuest – Dataset 1	.018	0.22	6.26	<.001
Dimension A – Dataset 2	.021	0.13	6.97	<.001
Dimension B Winsteps – Dataset 1	.024	2.32	8.60	<.001
Dimension B ConQuest – Dataset 1	.015	1.59	5.14	.002
Dimension B – Dataset 2	.024	1.25	8.20	<.001

Note. A comparison of R^2 results across software and data halves.

Since it was determined that the combination of the consecutive approaches, Dimension A and B, were the best fitting, canonical correlations were used to compare the multidimensional models by the Family Characteristics cluster. The canonical R^2 was used as an indicator. The combination of Dimension A and B for the second half of the data yielded a canonical R^2 = .027 (Table 28). The result was the largest canonical R^2 of the models examined, while still being comparable to the other values.

Table 28

Canonical R^2 Comparison of Dataset 1 and 2 by Family Characteristics

	Canonical R^2	F	Sig.
Combined Winsteps Dimension A & B – Dataset 1	.024	4.56	<.001
Combined ConQuest Dim A & B – Dataset 1	.019	3.91	.001
2-Dimensional ConQuest – Dataset 1	.023	4.15	<.001
Combined Dimension A & B – Dataset 2	.027	5.03	<.001

Note. Canonical comparisons across datasets and model types.

The item difficulty was compared between dataset 1 and 2 using the combination of the consecutive approach results. As Figure 22 showed, the item difficulty positions were very consistent with only one item, W8, showing substantial deviation.

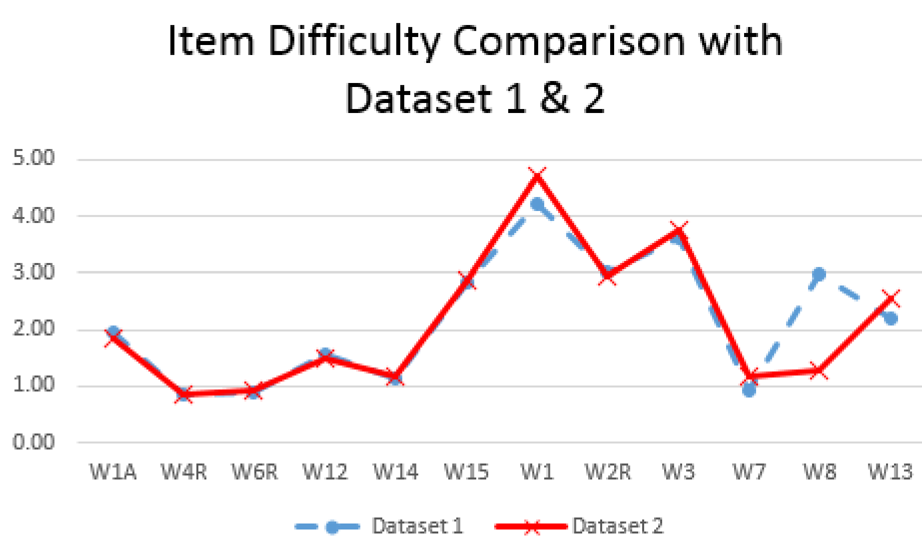


Figure 22. Item difficulty position across data halves.

Due to the structured skip protocols, the second half of the NSAP dataset had similar missing data issues as the first half. Table 29 demonstrates that the percentage of missing data was close for Dimension A but nearly double for Dimension B. The skewness and kurtosis indicated a lack of normality. Both dimensions were skewed with > 1.5 skewness. The kurtosis for these dimensions indicated that the data were not peaked, with values < 3.0, unlike Dimensions A and B for the first half of the data.

Table 29
Data Characteristics from Dataset 1 and Dataset 2

Model	Skewness	Kurtosis	Missing Data
Dataset 1 – Dimension A	1.86	3.69	58.8%
Dataset 2 – Dimension A	1.67	2.77	64.4%
Dataset 1 – Dimension B	2.02	3.79	32.6%
Dataset 2 – Dimension B	1.69	2.46	60.11%

Note. Additional results provided from ConQuest.

The second half of the data supported the model choice by providing consistent results with the first half of the dataset. The item fit and difficulty showed minor differences. The model fit indices were consistent with the first half of the data. The person logit testing resulted in scores near the first half of the data for both the multiple regression and the canonical correlations. The skewness and kurtosis were generally consistent between the data halves.

Model Selection Rationale

The intent of this study was to compare results using three approaches to examine the dimensionality and psychometric characteristics of a national dataset through the use of Rasch analyses. The unidimensional approach was used to explore the possibility that the data represented a single latent construct and used to establish a base model for comparison. Most often, the consecutive approach is used to estimate parameters for established dimensions within a multidimensional model. Also, if the data have been shown to fit a multidimensional model and no resources are available to the researcher to use a MIRT model, it has been recommended to use a consecutive approach to represent the data (Wiberg, 2012). For this study, the consecutive approach was used as an exploratory tool to draw to out dimensions. MIRT was then used to confirm the results of the consecutive approach findings compared to the unidimensional baseline.

During the study two major decisions were made that impacted the direction and findings of this research. The first was to examine the data for the possibility of more than one dimension within the data. The simplest, most direct model choice was the unidimensional model. The unidimensional model found that the first half of the data explained an adequate amount of variance at 56.4% with a sufficiently low first contrast eigenvalue. Based upon past psychometric research by Park, Barth, and Harrington (2013) and the use of the NSAP dataset by other researchers with only a single item or a cluster of items to support their conclusions (Malm, Vandivere, & Mcklindon, 2011; Radel, Bramlett, & Waters, 2010; Vandivere et al., 2009), multiple dimensions were examined. The consecutive approach was utilized to explore the possibility of additional dimensions. With this decision, lower eigenvalues were used as cutoffs. An examination of lower eigenvalues by Raiche in 2005 supported the use of adjusted eigenvalues based on sample size and number of items. An eigenvalue greater than 1.4 was used as the exploratory threshold. Eigenvalues of ≤ 1.4 were seen as the threshold for randomness (Smith & Miao, 1994). Utilizing this exploratory threshold, three dimensions were found through the consecutive approach and two MIRT models were tested, a 2-dimensional and 3-dimensional model.

The second impactful decision for this study was the best fitting model selection of the combined consecutive Dimensions A and B. It was determined that the 2-dimensional model, with appropriate infit and outfit mean squares, was a better fitting model than the unidimensional model through the use of AIC and G^2. With these results, the more common choice has been to select the 2-dimensional MIRT model with an adequate fit. If the consecutive approach is utilized to evaluate the dimensions within multidimensional model, the expectation has been that the MIRT will continue to be the best fitting model. The advantage to selecting a MIRT model that fits over the combining of two consecutive models, is that the MIRT model accounts for the possibility of the dimensions being interrelated. Additionally, the consecutive approach will produce an overestimated measurement error, since each separate model calculates an individual measurement error which are summed. The MIRT model represents a single measurement error calculation. Also, dimensions with the MIRT model are expected to correlate better than correlations between the dimensions/models of the consecutive approach (Allen & Wilson, 2006; Briggs & Wilson, 2003; Wiberg, 2012). Previous research has suggested that reliabilities of the unidimensional and MIRT models have been more consistent than consecutive approach models (Briggs & Wilson, 2003).

Unlike other studies, when comparing consecutive models to MIRT, this research found that the combined consecutive approach had a lower combined AIC than the MIRT model. AIC has been used to compare nonnested models (Akaike, 1985). The separation reliabilities were consistent between the MIRT and consecutive approaches. The standard errors were found to be most consistent and lower in the consecutive models A and B for both halves of the data. The largest error term was found in the 2-dimensional MIRT model. In addition, the error terms seemed to vary more in the ConQuest estimations (Figure 23). This variation supported the choice of the consecutive approach with a particular leaning towards the Winsteps software.

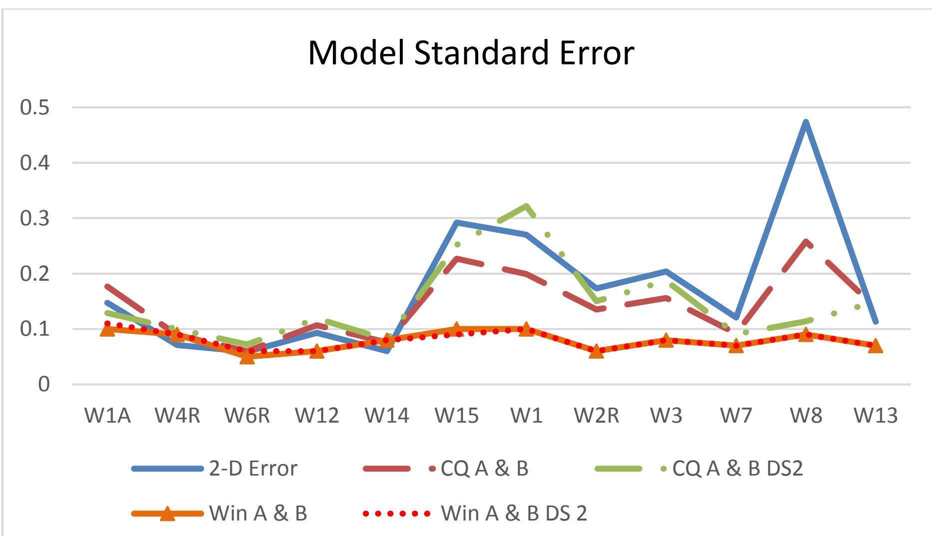


Figure 23. Standard error for items found in each model. Abbreviations for figure: ConQuest (CQ), Winsteps (Win), dataset 2 (DS2), and 2-dimensional model (2-D).

Another consideration for the consecutive approach was the use of Winsteps over ConQuest for this study. The original NSAP data collected used procedural skip logic, which opened up potential concerns regarding missing data. Winsteps utilizes joint maximum likelihood estimation (JMLE) to calculate estimation parameters and ConQuest utilizes marginal maximum likelihood estimation (MMLE). JMLE simultaneously calculates item and person estimates, while taking into account the rating scale, yielding consistent parameter estimates. Missing data impacts the sensitivity of the model fit but does not bias the parameter estimates in JMLE (Linacre, 2012). MMLE calculates the item estimates with the use of the rating scale and then estimates the person abilities. MMLE has a tendency of overestimating the item difficulty in comparison to the JMLE approach (Demars, 2002). The item difficulty estimates found in this study were consistently higher in the ConQuest estimates (Figure 24). ConQuest had greater variation for the item difficulty estimation for item W8 between dataset 1 and dataset 2. Item difficulty for Winsteps was consistently estimated with values below ConQuest estimations as well as consistently estimating item W8. These findings led to the final decision to use the combined consecutive models of Dimension A and B as best fitting the data.

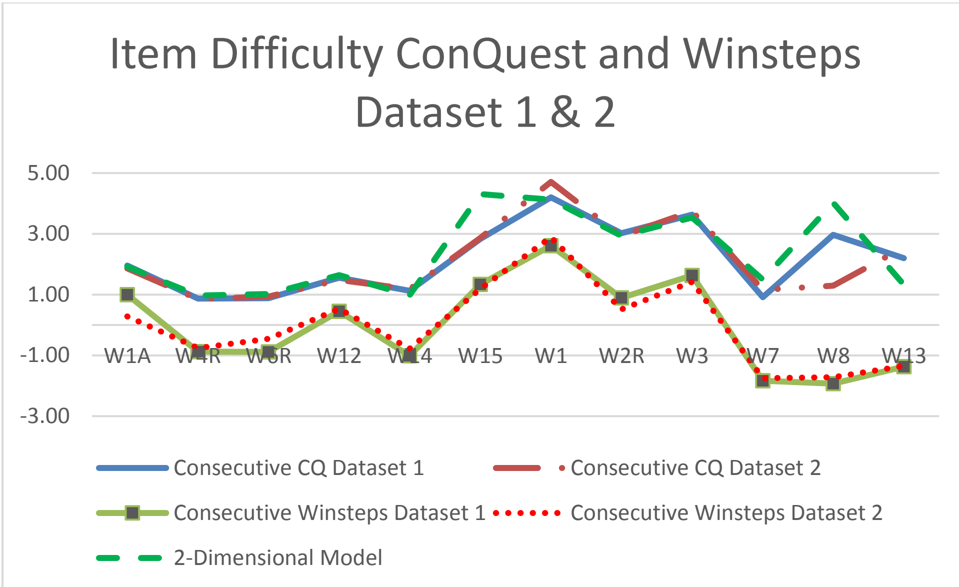


Figure 24. Item difficulty comparison between dataset halves and software.

Chapter Four: Discussion

This chapter summarizes the major findings of the analysis, according to the research question, discusses limitations of the study and provides recommendations for future research.

Major Findings by Research Question

Research question one.

1. Did the data from the parent-child well-being subsection found in the National Survey of Adoptive Parents (2007), NSAP, support a unidimensional or multidimensional structure when using a Rasch partial credit model for analysis?
    a. Were the psychometric properties of model fit, item fit, and reliability more suitable for the NSAP data within a unidimensional model or multidimensional model?
    b. How did item and person logit positions differ between the unidimensional and multi-dimensional findings?
    c. Did item and person indicators of position differ between software (Winsteps and ConQuest) when using the unidimensional and consecutive approaches?
    d. Which approach yielded a better model fit for the well-being subsection of NSAP?

It was valuable to determine if the data for the National Survey of Adoptive Parents fit a unidimensional or a multidimensional model in order to understand how to best utilizes the survey results. Since this was a national survey, the potential benefit to both adoption researcher and practitioners was invaluable, especially since the adoptive community has been relatively small and to have an opportunity for large-scale data collection about the community has been rare. Either result of unidimensional or multidimensional structure would have been beneficial to the adoptive community. A unidimensional model, focused on a single construct, makes the analysis simpler to interpret. For this study, the unidimensional model had 12 items, which resulted in a larger number of items per dimension than the multidimensional models found.

Since the NSAP was developed as an extension of the U.S. Census by the federal government, the development of the well-being subset was not the primary focus of the instrument, so testing to ensure that items supported a single construct was not done (Bramlett et al., 2010). Most often, the consecutive approach has been used for examination into an isolated dimension within an established multidimensional model, in order to provide isolated analysis on the dimension. In this study, the consecutive approach was used as an exploratory tool to determine and develop potential dimensions through Rasch analysis techniques. The resulting dimensions were further tested using a multidimensional approach. The multidimensional approach accounted for interrelationships between constructs that the consecutive approach is incapable of addressing.

As the NSAP data were analyzed it was accepted that multiple dimensions represented the data best. This finding was supported by past research that utilized multidimensional models or isolated items as representative constructs (Harwood et al., 2013; Lee, Yun, Yoo, & Nelson, 2010; Park et al., 2013). The best fitting model was selected from the consecutive approach, which combined two independent dimensions. As mentioned in the model selection rationale, it is unusual to select a consecutive approach model, when multiple dimensions are found, over a multidimensional model. Generally, the measurement error has been compounded through the use of the consecutive approach in contrast to a multidimensional approach. However, results of the multidimensional analysis may have been impacted by the amount of missing data as a possible reason for the higher measurement error found in the multidimensional model. The missing data was due to the skip protocols used during data collection, which suited the purposes of the involved agencies more than well-being and/or adoption research. These results of this study can serve to fill in the gap generated by the lack of focus in this subsections development and allows future researchers and practitioners to use the data from this survey more precisely in their analyses. Also, the findings will help to contribute insight into developing future adoption research surveys by providing a base of items for the identified constructs.

Research question two.

2. Were scale response categories used appropriately for each of the utilized items?

The majority of the items from the well-being subsection of the NSAP used for this analysis had clear Andrich Thresholds and needed no adjustment. However, two items, W4R and W14, used within the unidimensional, Dimension A, and the 2-dimensional models needed to be recalibrated. Recalibrating the items introduced more interpretable results and clarified the item difficulty by balancing the response categories. The ability to recalibrate the response scale of items is a strength of Rasch analysis. A calibrated scale yields with a consistent distance between scale scores, allowing for clarity in the interpretation of a scale score.

Research question three.

3. Once the dimensionality had been established and the item categories determined, did respondents for different adoption types (inter-country, domestic-private, and foster-to-adopt) interpret the items differently as observed through differential item functioning (DIF)? How did the differential item functioning results for adoption type compare between the unidimensional, the consecutive, and the multi-dimensional approach?

Differential item functioning was assessed to determine if adoptive parents participating in the survey responded differently to particular items. It was found that adoption type had a significant ($p \leq .01$) impact on responses to four items, which remained consistent across the models containing these items. Of the four items found to vary between the adoption types, two were found to be easier for inter-country adoptive parents to answer (W2R & W12), one was easier for foster-to-adopt parents to respond positively (W14), and one was easier for domestic-private adoptive parents to respond more positively to (W15). In general, very few of the responses to any of the items were negative, which indicated that when item variation was found between adoptive group types the group with the lower logit position was seen as the most positive in comparison to other groups. The items where the inter-country adoptive parents showed an easier time responding positively addressed the impact of the child on the family and the frequency of affection showed to the parent. It was found that foster-to-adopt parents responded more positively to the question that dealt with the child's feelings toward being adopted. Domestic-private adoption parents found it easier to respond positively to the question asking if they would repeat the adoption if they knew before the adoption everything they know now. It would be difficult to explain why these items were found to vary between the groups in the specific way identified without additional information.

Research question four.

4. How did the person logits and item difficulty compare for each dimension found within models from the unidimensional, consecutive, and multidimensional approaches compare across models?
    a. Did the person logit positions correlate across dimensions/models and software packages?
    b. Using a cluster of three independent variables, Adoptive Family with or without Biological Children, Child Lived with Birth Family, and Adoptive Parent/s and Child of Differing Races, as predictors in a regression analysis were the R^2 values comparable between the models and the software?

c. Using the same cluster of family characteristic variables as independent variable in a canonical correlation analysis, were the canonical R^2 values comparable between the models and the software?

d. How did the item difficulty compare across the models examined?

Inspecting person logit and item difficulty between the models and dimensions tested the agreement and consistency of the NSAP data. Some differences were revealed through these comparisons of logit positions and items difficulty between Winsteps and ConQuest. ConQuest consistently placed the item difficulty higher than Winsteps. The pattern of item difficulties was consistent across the models and software. The logit position comparison done through multiple regression and canonical correlations were found to be consistent between models and software with negligible differences in predictive capacity between models.

Research question five.

5. With the best fitting model selected, were the dimensionality and model fit replicated using the second half of the dataset for cross validation? Were the item fit, DIF, and validation measures comparable across the two halves of the dataset?

Splitting the NSAP dataset in half allowed for a comparison of the findings. The expectation was that a comparison of the results would be consistent across the two dataset halves. Care was taken to balance the two halves of data to ensure that the identified groups were represented in each half and to address the data collection issues that were driven by the age of selected child. Overall the second half of the dataset fit similarly to the data as found when analyzing the first half of the dataset. Consistent with the first half of the dataset, the item positions were comparable across the approaches. The convergence in ConQuest was more difficult in the second half of the data, possibly due to the location of the missing data within the dataset. Re-ordering of the data file helped ConQuest to render results, during the second half of the study. The R^2 values, used for validity, were small which limited the ability of the analysis to validate the results. The decision to select the consecutive approach was based on model fit with little additional supporting evidence.

Survey Development

According to the National Center for Health Statistics of the U.S. Department of Health and Human Services the development of the NSAP survey had at least five agencies or groups involved with its creation. The survey developers used questions from existing surveys and also developed some new questions. The purpose was to collect data regarding the adoption community with a primary focus on the foster-to-adopt community (Bramlett, Foster, et al., 2010). The focus on foster-to-adopt community was due to the local and federal government’s financial investment and direct oversight on the foster care system when compared to the other forms of adoption. The NSAP was used to gather data to provide insight into a broad variety of topics, while using a national sample.

Creating a survey with at least five contributing groups, with competing understandings and expectations, is a difficult undertaking and that was how the NSAP survey was created. Expert insight into the development of a survey is invaluable; however, if the expert contributions muddle the purpose, advice on content may be confused. Evidence of competing purposes was seen throughout the variety of topics addressed through the questions within the survey. Some questions from the survey tapped into perceptions of how the parent felt the experience of adoption went for them, their spouse, and their child. Other questions addressed perception of the relationship between members of the family, while other questions were asked about government resources used by the families. These may all be valuable questions but they do not contribute necessarily to the construct of well-being. With multiple groups utilized for the development of this instrument, it would be expected to have differing interpretations of the well-being focus guiding the groups' item selections.

Since the NSAP was the largest nation-wide data collection endeavor for the adoptive community undertaken, the potential for gathering of useful parent-child well-being data was vast. However, an unfocused governmental approach caused the data collection opportunity to fall short of its potential. Had this survey been developed with a small number of primary contributors the focus of the tool as well as the model fit might have been better. A partnership between a content expert and a psychometrician when developing this tool would have maintained a narrower objective than the five contributing groups with different objectives. The partnership would have been able to pilot the items through a psychometric lens and not just a survey protocol perspective. The focused development and earlier testing would have improved the overall value and usefulness of NSAP.

Summary of Results Regarding Value of the NSAP

The Rasch analysis of the well-being subsection of the NSAP has found that psychometrics of the survey matched the process of its creation. Over half of the items were eliminated from the analysis due to specified skip logic. The person logit positions indicated that the majority of the questions in this survey were (too) easy for the participants to agree with, which indicated that items did not draw out the nuances of the latent variable, well-being. The survey lacked the sensitivity and consistency to distinguish between persons of differing opinions. Since so few parents indicated negative responses, it could have been possible that the parents were responding positively to the person collecting the data on the phone due to response bias, answering the questions to please the interviewer and to avoid embarrassing themselves. The findings of this analysis provided insight into the measure dimensionality but did not completely resolve the dimensionality questions. Narrowing the focus of the survey and topic area of the survey would have improved the overall influence of the findings.

Summary of Results Regarding the Three Approaches

Three approaches (unidimensional, consecutive, and MIRT) were taken to examine the NSAP data to determine the best fitting model. A unidimensional approach is taken when there is one known dimension or as the initial exploration into the data. Rasch analysis operates under the assumption that the data are describing a single construct and whenever possible keeping this singular focus is advantageous (Linacre, 1998). Having a single construct described by the model allows for easier understanding of the model fit, item difficulty and fit, scale calibration, and person ability (H. L. Chang & Shih, 2012).

The multidimensional item response theory approach takes the interrelationship between the dimensions into account when examining the model fit. The MIRT approach is used when you have a known multidimensional model. The relationships between the dimensions can be inspected and adjusted to find the best fitting model. Items fit and scale calibration can be adjusted within this model but the interpretation of the results become murkier due to the complexity of the model design. Although more complicated to interpret the MIRT model allows for more complexity within the data than the unidimensional and consecutive approaches (Briggs & Wilson, 2003; H. L. Chang & Shih, 2012).

The consecutive approach is utilized for two primary reasons. The first reason this approach is used occurs when a researcher needs to examine a single dimension within preexisting multidimensional model. This approach allows for an investigation of the fit statistics of the dimension without the interrelationship of the other dimensions interfering with the results. This type of investigation might be valuable to understand the strength of each dimension and may help to identify weaker items within the isolated dimension. The consecutive approach can also be used as an exploratory tool when the dimensionality is not fully known. Utilizing the consecutive approach employs a step by step inspection of the items to determine if there is enough connection for the items to form a separate dimension. This approach does not account for or determine the relationship between the dimensions, which supports isolated dimensional interpretations.

These dimensional interpretations are the same as the unidimensional approach with a single construct for each dimension (Purya Baghaei, 2013; Wiberg, 2012).

When investigating the dimensionality through the consecutive approach, a second dimension should be considered if past research has indicated multiple dimensions or when the analysis indicated the possibility of misfitting items forming another dimension from the unexplained variance. Additionally, items trimmed from the unidimensional model can be analyzed as a possible dimension, assuming that there are at least three items. From a practical perspective, it is important for there to be adequate interest in pursuing a second dimension, otherwise, the unidimensional model is preferred (J M Linacre, 1998).

Selecting the best fitting model looked to the parsimony of the model and the goodness of fit. Parsimony is achieved with the fewest number of parameter estimates and the simplest explanation. For this research AIC was used as the goodness of fit indicator. The final determination inspected the 12 most consistent items by first comparing the nested models (unidimensional and 2-dimensional model). The 2-dimensional model was a better fit than the unidimensional model. Next, the nonnested models (the consecutive approach models and the 2-dimensional model) were compared and the consecutive model was selected with a better AIC fit better and 1less parameter estimate than the 2-dimensional model. Selecting the consecutive approach has allowed for the two dimensions to be examined and interpreted separately.

Despite the selection of the consecutive approach for this study, the two primary approaches that I recommend when developing an IRT model are the unidimensional and multidimensional approaches. The unidimensional model is the foundation of Rasch as well as the starting point for the consecutive approach. When the dimensionality of the data is unknown, begin with the Rasch unidimensional approach. Starting with the Rasch approach allows simplest, most interpretable results, and if there are additional dimensions then the consecutive approach can be used. When the dimensionality is known and there is more than one dimension, the multidimensional approach is generally the best approach to use. MIRT accounts for the relationships between the dimensions as well as reduces the error terms. The consecutive approach can be used to explore the dimensions within a MIRT model, utilizing a unidimensional approach for each item set, when necessary. The consecutive approach is a supporting analysis technique and not a primary tool.

Limitations

The survey was given to only one adoptive parent and not both parents, if applicable for each selected child, which reveals only one perspective of the examined relationship. In addition, the child was not given an opportunity to express their perspective on the parent-child relationship. Only one child was identified within each home, so those homes with more than one adopted child needed to respond only for the selected child. It was also assumed that the adoptive parent who had more than one child was able provide isolated responses and not responses that represented a composite of all of the children to the questions.

The person logit position suggest that the participants were highly agreeable, which might be explained by most participants have a strong desire to describe the parent-child relationship as being positive. The participants may have also seen positive responses as the expected or preferred responses. These tendencies by the participants could have impacted their selection patterns.

As noted earlier, the survey logic determined skip patterns that resulted in large amounts of missing data. The large amount of missing data impacted the use of the majority of items within the Well-Being subset of the NSAP dataset and may have reduced the accuracy of the results of the multidimensional analysis.

The selection of the consecutive approach model was not validated through the validity analyses. The validation analysis yielded minute comparison variances, which left the model selection unsubstantiated.

Recommendations for Further Study

Through the guidance of a psychometrician and an adoption or well-being content expert, it is recommended that the current results be treated as a pilot study and the survey enhanced through adjustments to the items. It is recommended that the revised well-being section of the NSAP survey be re-administered with another population, removing the procedural skip logic, which limited the data collected by all participants. This would decrease the overall missing data as well as provide a more complete understanding of the results. More items added to the survey would broaden the potential to explain the latent constructs of the child's interaction with others and the parent's expectations. Future research should also pay attention to the spread of item difficulty for these constructs, particularly the child's interaction with others, due to the tendency of the items in this dimension to be quite easy to agree with. Items W8, W4R, and W14 should also be explored to determine if these items should be discarded. Finally, the next step of this research would be to validate the findings of this survey with one or more measures of associated constructs related to the dimensions of parent-child relational well-being.

It is recommended that future research be conducted to continue to test the use of the consecutive approach as an exploratory tool. Determining if the consecutive approach aids in the investigation of possible multidimensional models when there are large proportions of missing data as well as its potential as a tool for surveys with unknown dimensionality would be useful for applied researchers. Compare the proposed dimensionality using IRT consecutive approach as an exploratory tool with results from a classical test theory approach such as principle component analysis or exploratory factor analysis. As dimensionality is examined using the results of past dimensionality findings should be considered as either a starting point or as evidence of potential dimensionality. When using the consecutive approach as an exploratory tool, starting with stricter infit and outfit cutoff criteria and relaxing the criteria, within accepted bounds, can suggest the level of confidence in the found dimensions. For this study, past analysis of the data suggested multiple dimensions; however, the stricter criteria suggested a unidimensional model. Loosening the criteria yielded more than one dimension, though the resulting analysis was not as clean as desired. Further study of the standardized residuals, through a principal components analysis of residuals, provides insight into the dimensionality of the data. The use of item residual inter-correlations can reveal potential dimensions through the correlation values clustering between specific items (Linacre, 1998). Future research into the psychometrics of the measure could help to provide a clearer result.

Since the nonnested model was selected in this study, additional model testing could be done by utilizing comparisons designed to examine nonnested models besides the AIC. When comparing nonnested models with normally distributed or flat data distributions, the Vuong test can be used for comparisons. If the data suggest a skewed or peaked distribution, then Clarke (2009) has developed a distribution-free comparison test that can be used.

Improvements to the validation measures is recommended to identify greater differences between the models in order and aid in model selection. Selecting different family characteristics might be a way to improve the findings. Another improvement would be to find another measure that addressees the well-being of the parent-child relationship and administering it alongside the studied measure. Correlations of the results could be compared to test the validity.

Besides adjusting how the data were collected, a future study could treat the missing data as missing not at random (MNAR) versus ignoring the missing data (Liu & Wang, 2017; Mariel & Enciso, 2016). In this future study, the researcher would need to impute the missing data in order to have a complete dataset. It is likely that fewer items would have been removed from the analysis as well as the results rendered from ConQuest improved. Simulations could be run to compare the parameter estimates produced by Winsteps and ConQuest to further compare the algorithm parameter estimation similarities and differences.

References

Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice*, *22*, 37–51. http://doi.org/10.1111/j.1745-3992.2003.tb00136.x

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. http://doi.org/10.1109/TAC.1974.1100705

Akaike, H. (1985). Prediction and entropy. In A. C. Atkinson & S. E. Fienberg (Eds.), *A Celebration of Statistics* (pp. 387–410). New York, NY: Springer New York. http://doi.org/10.1007/978-1-4613-8560-8_1

Allen, D. D., & Wilson, M. (2006). Introducing multidimensional item response modeling in health behavior and health education research. *Health Education Research*, *21*(SUPPL.1), 73–84. http://doi.org/10.1093/her/cyl086

Andrich, D. (1978). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement*, *2*(4), 581–594. http://doi.org/10.1177/014662167800200413

Baghaei, P. (2013). Development and psychometric evaluation of a multidimensional scale of willingness to communicate in a foreign language. *European Journal of Psychology of Education*, *28*(3), 1087–1103. http://doi.org/10.1007/s10212-012-0157-y

Baghaei, P., & Grotjahn, R. (2014). Establishing the construct validity of conversational C-Tests using a multidimensional Rasch model. *Psychological Test and Assessment*

Modeling, *56*(1), 60–82. Retrieved from http://www.psychologie-aktuell.com/fileadmin/download/ptam/1-2014_20140324/04_Baghaei.pdf

Bahraini, N. H. (2008). *Development and Validation of the Professional Support Inventory for Mental Health Professionals (Doctoral Dissertation)*. University of Denver.

Beijersbergen, M. D., Juffer, F., Bakermans-Kranenburg, M. J., & van Ijzendoorn, M. H. (2012). Remaining or becoming secure: Parental sensitive support predicts attachment continuity from infancy to adolescence in a longitudinal adoption study. *Developmental Psychology*, *48*(5), 1277–1282. http://doi.org/10.1037/a0027442

Berrick, J. D. (2008). *Take me home: Protecting America's vulnerable children and families*. New York, New York: Oxford University Press.

Berry, M. (1991). The practice of open adoption: Findings from a study of 1396 adoptive families. *Children and Youth Services Review*, *13*(5–6), 379–395. http://doi.org/10.1016/0190-7409(91)90027-F

Berry, M. (1998). Adoption in an era of family preservation. *Children and Youth Services Review*, *20*(1–2), 1–12. http://doi.org/10.1016/S0190-7409(97)00064-9

Berry, M., Dylla, D. J. C., Barth, R. P., & Needell, B. (1998). The role of open adoption in the adjustment of adopted children and their families. *Children and Youth Services Review*, *20*(1–2), 151–171. http://doi.org/10.1016/S0190-7409(97)00071-6

Bolsinova, M., & Maris, G. (2016). Can IRT solve the missing data problem in test equating? *Frontiers in Psychology*, *6*(JAN), 1–13. http://doi.org/10.3389/fpsyg.2015.01956

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (Second). New York: Routledge.

Bramlett, M. D., Brooks, K. S., Foster, E. B., Satorius, J., Frasier, A. M., Skalland, B. J., … Chowdhury, S. R. (2010). Design and operation of the National Survey of Adoptive Parents of Children With Special Health Care Needs, 2008. *Vital and Health Statistics. Ser. 1, Programs and Collection Procedures*, (51), 1–118. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20629282

Bramlett, M. D., Foster, E. B., Frasier, A. M., Satorius, J., Skalland, B. J., Nysse-Carris, K. L., … Chowdhury, S. R. (2010). Design and operation of the National Survey of Adoptive Parents, 2007. *Vital and Health Statistics. Ser. 1, Programs and Collection Procedures*, (50), 1–154. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20737837

Briggs, D. C., & Wilson, M. (2003). An Introduction to Multidimensional Measurement using Rasch Models. *Journal of Applied Measurement*, *4*(1), 87–100.

Brodzinsky, C. D. M., Smith, D. W., & Brodzinsky, A. B. (1998). *Children's Adjustment to Adoption: Developmental and Clinical Issues* (38th ed.). Thousand Oaks: Sage Publications, Inc. http://doi.org/10.4135/9781452220581

Burke, R. V., Schlueter, C., Vandercoy, J., & Authier, K. J. (2014). Post-Adoption services for families at risk of dissolution: A case study describing two families' experiences. *Clinical Case Studies*. http://doi.org/10.1177/1534650114556696

Carnes-Holt, K., & Bratton, S. C. (2014). The efficacy of child parent relationship therapy for adopted children with attachment disruptions. *Journal of Counseling and*

Development, *92*(3), 328–337. http://doi.org/10.1002/j.1556-6676.2014.00160.x

Chang, C.C., Su, J.A., Tsai, C.S., Yen, C.F., Liu, J.H., & Lin, C.Y. (2015). Rasch analysis suggested three unidimensional domains for Affiliate Stigma Scale: Additional psychometric evaluation. *Journal of Clinical Epidemiology*, *68*(6), 674–683. http://doi.org/10.1016/j.jclinepi.2015.01.018

Chang, H. L., & Shih, C. K. (2012). Using a multidimensional Rasch model approach to measure the police's perceived ability to detect, detain and intercept DWI vehicles when conducting sobriety checkpoints. *Accident Analysis and Prevention*, *48*, 505–517. http://doi.org/10.1016/j.aap.2012.03.012

Cheng, Y.-Y., Wang, W.-C., & Ho, Y.-H. (2009). Multidimensional Rasch Analysis of a psychological test with multiple subtests: A statistical solution for the Bandwidth--Fidelity Dilemma. *Educational and Psychological Measurement*, *69*(3), 369–388. http://doi.org/10.1177/0013164408323241

Child Welfare Information Gateway. (2013). *Foster Care Statistics 2013* (Vol. 4, Part 13). Retrieved from http://www.gpo.gov/fdsys/pkg/CFR-2012-title45-vol4/pdf/CFR-2012-title45-vol4-part1355.pdf

Clarke, K. A. (2009). A Simple distribution-free test for nonnested model selection. *Ssrn*, 347–363. http://doi.org/10.1093/pan/mpm004

Colonnesi, C., Wissink, I. B., Noom, M. J., Asscher, J. J., Hoeve, M., Stams, G. J. J. M., … Kellaert-Knol, M. G. (2013). Basic trust: An attachment-oriented intervention based on mind-mindedness in adoptive families. *Research on Social Work Practice*, *23*(2), 179–188. http://doi.org/10.1177/1049731512469301

Demars, C. (2002). Applied measurement in education incomplete data and item parameter estimates under JMLE and MML estimation incomplete data and item parameter estimates under JMLE and MML estimation. *Applied Measurement in Education*, *15*(1), 15–31. http://doi.org/10.1207/S15324818AME1501

DeMars, C. (2002). Missing data and IRT item parameter estimation. In *Annual Meeting of the American Educational Research* (p. 15). Retrieved from http://www.jmu.edu/assessment/wm_library/missdata.pdf

Engel, M., Phillips, N. K., & Dellacava, F. a. (2007). International adoption: A sociological account of the US experience. *International Journal of Sociology and Social Policy*, *27*(5/6), 257–270. http://doi.org/10.1108/01443330710757285

Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/person statistics. *Educational and Psychological Measurement*, *58*(3), 357–381. http://doi.org/10.1177/0013164498058003001

Gilbert, N., Parton, N., & Skivenes, M. (2011). *Child protection systems: International trends and orientations*. New York: Oxford University Press.

Goldberg, S., Muir, R., & Kerr, J. (Eds.). (1995). *Attachment theory: Social, developmental, and clinical perspectives*. Hillsdale, New Jersey: The Analytis Press, Inc.

Goodman, G. S., Emery, R. E., & Haugaard, J. J. (1998). Developmental psychology and law: Divorce, child maltreatment, foster care and adoption. In K. A. Damon, William; Sigel, Irving E.; Renninger (Ed.), *Handbook of child psychology* (5th ed., pp. 775–874). Hoboken, NJ: John Wiley & Sons Inc.

Grotevant, H. D., Ross, N. M., Marchel, M. a., & Mcroy, R. G. (1999). Adaptive behavior in adopted children: Predictors from early risk, collaboration in relationships within the adoptive kinship network, and openness arrangements. *Journal of Adolescent Research*, *14*(2), 231–247. http://doi.org/10.1177/0743558499142005

Hartinger-Saunders, R. M., Trouteaud, A., & Matos-Johnson, J. (2014). Post adoption service need and use as predictors of adoption dissolution: Findings from the 2012 National Adoptive Families Study. *Adoption Quarterly*, *6755*(July 2015), 150527101018004. http://doi.org/10.1080/10926755.2014.895469

Harwood, R., Feng, X., & Yu, S. (2013). Preadoption adversities and postadoption mediators of mental health and school outcomes among international, foster, and private adoptees in the United States. *Journal of Family Psychology : JFP : Journal of the Division of Family Psychology of the American Psychological Association (Division 43)*, *27*(3), 409–20. http://doi.org/10.1037/a0032908

Hjern, A., Lindblad, F., & Vinnerljung, B. (2002). Suicide, psychiatric illness, and social maladjustment in intercountry adoptees in Sweden: A cohort study. *Lancet*, *360*(9331), 443–448. http://doi.org/10.1016/S0140-6736(02)09674-5

Hollinger, J. H. (2000). *Adoption law and practice* (Vol. 1). New York: Matthew Bender. Retrieved from https://books.google.com/books?id=qds7AQAAIAAJ

Huang, H. Y., Wang, W. C., Chen, P. H., & Su, C. M. (2013). Higher-order item response models for hierarchical latent traits. *Applied Psychological Measurement*, *37*(8), 619–637. http://doi.org/10.1177/0146621613488819

Judge, S. (2004). The impact of early institutionalization on child and family outcomes. *Adoption Quarterly, 7*(3), 31–48. http://doi.org/10.1300/J145v07n03_02

Juffer, F., & van Ijzendoorn, M. H. (2005). Behavior problems and mental health referrals of international adoptees: A meta-analysis. *JAMA : The Journal of the American Medical Association*, *293*(20), 2501–15. http://doi.org/10.1001/jama.293.20.2501

Lee, R. M. (2003). The transracial adoption paradox: History, research, and counseling implications of cultural socialization. *The Counseling Psychologist*, *31*(6), 711–744. http://doi.org/10.1177/0011000003258087

Lee, R. M., Yun, A. B., Yoo, H. C., & Nelson, K. P. (2010). Comparing the ethnic identity and well-being of adopted Korean Americans with immigrant/US-born Korean Americans and Korean international students. *Adoption Quarterly*, *13*(1), 2–17. http://doi.org/10.1080/10926751003704408

Linacre, J. M. (1998). Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement*, *2*(3), 266–83. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9711024

Linacre, J. M. (2002). Review of Reviews of Bond & Fox (2001). *Transactions of the Rasch Measurement SIG Autumn 2002*, *16*(2), 871–882.

Linacre, J. M. (2012). *Winsteps help for Rasch analysis: Rasch-model computer programs*. http://doi.org/ISBN 0-941938-03-4

Liu, C. W., & Wang, W. C. (2017). Non-ignorable missingness item response theory models for choice effects in examinee-selected items. *British Journal of*

Mathematical and Statistical Psychology, *70*(3), 499–524. http://doi.org/10.1111/bmsp.12097

Malm, K., Vandivere, S., & Mcklindon, A. (2011). *Children Adopted from Foster Care : Adoption Agreements, Adoption Subsidies, and Other Post - Adoption Supports*. (Is this a book? If so, it's missing publisher, etc.)

Mariel, S., & Enciso, S. (2016). the Effects of Missing Data Treatment on Person Ability Estimates Using Irt Models.(Is this a book?)

Miller, B. C., Fan, X., & Grotevant, H. D. (2005). Methodological issues in using large-scale survey data for adoption research. In D. Brodzinsky & J. Palacios (Eds.), *Psychological issues in adoption: Research and practice* (pp. 233–256). Westport, Connecticut: Praeger Publishers.

Mountjoy, T. P., & Vanlandingham, E. N. (2015). *Effects of attachment styles of foster and adoptive parents on the relational interactions of their foster and adoptive children (Masters Thesis)*. California State Univeristy, San Bernardino, CA. Retrieved from http://scholarworks.lib.csusb.edu/etd

O'Neill, M., Risley-Curtiss, C., Ayón, C., & Williams, L. R. (2012). Placement stability in the context of child development. *Children and Youth Services Review, 34*(7), 1251–1258. http://doi.org/10.1016/j.childyouth.2012.02.018

Pace, C. S., & Zavattini, G. C. (2011). "Adoption and attachment theory" the attachment models of adoptive mothers and the revision of attachment patterns of their late-adopted children. *Child: Care, Health and Development*, *37*(1), 82–88. http://doi.org/10.1111/j.1365-2214.2010.01135.x

Palacios, J., & Sanchez-Sandoval, Y. (2005). Beyond adopted/nonadopted comparisons. In D. Brodzinsky & J. Palacios (Eds.), *Psychological issues in adoption: Research and practice* (pp. 117–144). Westport, Connecticut: Praeger Publishers.

Park, H., Barth, R. P., & Harrington, D. (2013). Factor structure of adoptive parent-child relationship items from the National Study of Adoptive Parents. *Journal of the Society for Social Work and Research*, *4*(1), 20–30. http://doi.org/10.5243/jsswr.2013.2

Radel, L. F., Bramlett, M. D., & Waters, A. (2010). Legal and informal adoption by relatives in the US: Comparative characteristics and well-being from a nationally representative sample. *Adoption Quarterly*, *13*(3–4), 268–291. http://doi.org/10.1080/10926755.2010.537957

Rahayah Ariffin, S., Omar, B., Isa, A., & Sharif, S. (2010). Validity and reliability multiple intelligent item using Rasch measurement model. *Procedia - Social and Behavioral Sciences*, *9*, 729–733. http://doi.org/10.1016/j.sbspro.2010.12.225

Raiche, G. (2005). Critical eigenvalue sizes (variances) in standardized residual principal components analysis. *Rasch Measurement Transactions*, *19*(1), 1012. Retrieved from https://www.rasch.org/rmt/rmt191h.htm

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests.* (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press.

Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer.

http://doi.org/10.1007/978-0-387-89976-3

Rees, C. (2008). Children's attachments. *Paediatrics and Child Health*, *18*(5), 8. http://doi.org/10.1016/j.paed.2008.02.009

Rost, J., & Carstensen, C. H. (2002). Multidimensional Rasch measurement via item component models and faceted designs. *Applied Psychological Measurement*, *26*(1), 42–56. http://doi.org/10.1177/0146621602026001003

Royal, K. D., Ellis, A., Ensslen, A., & Homan, A. (2010). Rating scale optimiazation in survey research: An application of the Rasch rating scale model. *Journal of Applied Quantitative Methods*, *5*, 607–617.

Rutter, M. (2005). Adverse preadoption experiences and psychological outcomes. In D. Brodzinsky & J. Palacios (Eds.), *Psychological issues in adoption: Research and practice* (pp. 67–92). Westport, Connecticut: Praeger Publishers.

Sass, A. T. L. (2014). *Racial identity development of white parents of transracial adoptees : A narrative approach* (Doctoral dissertation). Retrieved from http://hdl.handle.net/2345/3679

Sharkness, J. (2014). Item response theory: Overview, applications, and promise for institutional research. *New Directions for Institutional Research*, *161*, 18. http://doi.org/10.1002/ir

Sharkness, J., & DeAngelo, L. (2011). Measuring student involvement: A comparison of classical test theory and item response theory in the construction of scales from student surveys. *Research in Higher Education*, *52*(5), 480–507. http://doi.org/10.1007/s11162-010-9202-3

Simon, R. J., & Altstein, H. (2000). *Adoption across borders: Serving the children in transracial and intercountry adoptions*. Lanham, Maryland: Rowman & Littlefield.

Singer, E., & Krebs, M. (2008). Children adopted at older ages, *34*(2), 170–174. (missing the journal title?)

Skivenes, M., & Tefre, Ø. S. (2012). Adoption in the child welfare system - A cross-country analysis of child welfare workers' recommendations for or against adoption. *Children and Youth Services Review*, *34*(11), 2220–2228. http://doi.org/10.1016/j.childyouth.2012.07.013

Smith, R. M., & Miao, C. Y. (1994). Assessing unidimensionality for Rasch measurement. In L. ABC-CLIO (Ed.), *Objective Measurement: Theory into Practice* (p. 338). Objective Measurement: Theory into Practice. (What is this last part of the citation?)

Sroufe, L. A., Egeland, B., Carlson, E. A., & Collins, W. A. (2009). *The development of the person: The Minnesota study of risk and adaptation from birth to adulthood*. New York, NY: Guilford Press.

Stams, G. J., Juffer, F., Rispens, J., & Hoksbergen, R. a. (what's this "a" doing?) (2000). The development and adjustment of 7-year-old children adopted in infancy. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *41*(8), 1025–1037. http://doi.org/10.1111/1469-7610.00690

Steele, M., Hodges, J., Kaniuk, J., Hillman, S., & Henderson, K. (2003). Attachment representations and adoption: Associations between maternal states of mind and emotion narratives in previously maltreated children. *Journal of Child*

Psychotherapy, *29*(2), 187–205. http://doi.org/10.1080/0075417031000138442

Tan, T. X., Major, D., Marn, T., Na, E., & Jackson, A. L. (2015). Adopted children's country of origin and post-adoption parent–child relationship quality: Findings from the United States National Survey of Adoptive Parents (NSAP). *Children and Youth Services Review*, *48*, 117–125. http://doi.org/10.1016/j.childyouth.2014.12.001

Tieman, W., van der Ende, J., & Verhulst, F. C. (2005). Psychiatric disorders in young adult intercountry adoptees: An epidemiological study. *The American Journal of Psychiatry*, *162*(3), 592–598. http://doi.org/10.1176/appi.ajp.162.3.592

Tieman, W., van der Ende, J., & Verhulst, F. C. (2006). Social functioning of young adult intercountry adoptees compared to nonadoptees. *Social Psychiatry and Psychiatric Epidemiology*, *41*(1), 68–74. http://doi.org/10.1007/s00127-005-0995-x (and this 'x'?)

van Ijzendoorn, M. H., Goldberg, S., Kroonenberg, P. M., & Frenkel, O. J. (1992). The relative effects of maternal and child problems on the quality of attachment: A meta-analysis of attachment in clinical samples. *Child Development*, *63*(4), 840–858. http://doi.org/10.2307/1131237

van Ijzendoorn, M. H., & Juffer, F. (2006). The Emanuel Miller Memorial Lecture 2006: Adoption as intervention. Meta-analytic evidence for massive catch-up and plasticity in physical, socio-emotional, and cognitive development. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *47*(12), 1228–1245. http://doi.org/10.1111/j.1469-7610.2006.01675.x (is the "x" part of the url?)

van Ijzendoorn, M. H., Juffer, F., & Poelhuis, C. W. K. (2005). Adoption and cognitive development: A meta-analytic comparison of adopted and nonadopted children’s IQ and school performance. *Psychological Bulletin*, *131*(2), 301–16. http://doi.org/10.1037/0033-2909.131.2.301

Vandivere, S., Malm, K., & Radel, L. F. (2009). *Adoption USA: A chartbook based on the 2007 National Survey of Adoptive Parents*. Washington, DC: U.S. Department of Health and Human Services; Office of the Assistant Secretary for Planning and Evaluation.

Vandivere, S., & McKlindon, A. (2010). The well-being of US children adopted from foster care, privately from the United States and internationally. *Adoption Quarterly*, *13*(3–4), 157–184. http://doi.org/10.1080/10926755.2010.524871

Verhulst, F. C., Althaus, M., & Versluis-Den Bieman, H. J. M. (1990). Problem behavior in international adoptees: II. Age at placement. *Journal of the American Academy of Child & Adolescent Psychiatry*, *29*(1), 104–111.

Viana, A. G., & Welsh, J. a. (should this "a" be capitaliized?) (2010). Correlates and predictors of parenting stress among internationally adopting mothers: A longitudinal investigation. *International Journal of Behavioral Development*, *34*(4), 363–373. http://doi.org/10.1177/0165025409339403

Vorria, P., Papaligoura, Z., Sarafidou, J., Kopakaki, M., Dunn, J., van Ijzendoorn, M. H., & Kontopoulou, A. (2006). The development of adopted children after institutional care: A follow-up study. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *47*(12), 1246–1253. http://doi.org/10.1111/j.1469-7610.2006.01666.x (is the "x" part of the url?)

Wiberg, M. (2012). Can a multidimensional test be evaluated with unidimensional item response theory?. *Educational Research and Evaluation*, *18*(4), 307–320. http://doi.org/10.1080/13803611.2012.670416

Wolfgram, S. M. (2008). Openness in adoption: What we know so far: A critical review of the literature. *The Social Worker*, *53*(2), 133–142.

Woodhouse, S. S., Dykas, M. J., & Cassidy, J. (2009). Perceptions of secure base provision within the family. *Attachment & Human Development*, *11*(1), 47–67. http://doi.org/10.1080/14616730802500792

Woolgar, M., & Scott, S. (2013). The negative consequences of over-diagnosing attachment disorders in adopted children: The importance of comprehensive formulations. *Clinical Child Psychology and Psychiatry*, *19*(3), 355–366. http://doi.org/10.1177/1359104513478545

Yu, C. H., Popp, S. O., Digangi, S., & Jannasch-Pennell, A. (2007). Assessing unidimensionality : A comparison of Rasch Modeling, parallel analysis , and TETRAD. *Practical Assessment, Research, and Evaluation*, *12*(14), 1–19.

Zamostny, K. P., O'Brien, K. M., Baden, A. L., & Wiley, M. O. (2003). The Practice of Adoption: History, Trends, and Social Context. *The Counseling Psychologist*, *31*(6), 651–678. http://doi.org/10.1177/0011000003258061

Zill, N. (2011). *Children and families adoption from foster care: Aiding children while saving public money*. Washington, DC: Brookings.

Zill, N., & Bramlett, M. D. (2014). Health and well-being of children adopted from foster care. *Children and Youth Services Review, 40*, 29–40. http://doi.org/10.1016/j.childyouth.2014.02.008

Appendix A – Well-being Subset of NSAP Survey Items

Item List
[S.C] = selected child
W1 – How would you describe your relationship to [S.C.]?

(1) Very Warm and Close
(2) Somewhat Warm and Close
(3) Somewhat Distant
(4) Very Distant

W1A – How would you describe your [spouse's/partner's] relationship to [S.C.]?

(1) Very Warm and Close
(2) Somewhat Warm and Close
(3) Somewhat Distant
(4) Very Distant

W2R – How often is [S.C.] affectionate or tender with you?

(1) Always
(2) Usually
(3) Sometimes
(4) Rarely
(5) Never

W3 – How satisfied are you with how affectionate or tender [S.C.] is with you?

(1) Very Satisfied
(2) Somewhat Satisfied
(3) Somewhat Dissatisfied
(4) Very Dissatisfied

W4R – Do you feel that [S.C.] and you make decisions about [his/her] life together?

(1) Always
(2) Usually
(3) Sometimes
(4) Rarely
(5) Never

W5 – During the past month, how often have you felt that you just did not understand [him/her]?

(1) Never
(2) Rarely
(3) Sometimes
(4) Usually
(5) Always

W6R – During the past month, how often have you felt that you can really trust [him/her]?

(1) Always
(2) Usually
(3) Sometimes
(4) Rarely
(5) Never

W7 – Thinking about [S.C.]'s relationship with you, would you say things are…?

(1) Better than you ever expected
(2) About what you expected
(3) More difficult than you ever expected

W8 – Thinking about [S.C.]'s relationship with your [spouse/partner], would you say things are…?

(1) Better than you ever expected
(2) About what you expected
(3) More difficult than you ever expected

W9 – How often does [S.C.] experience difficulty in getting along with other children in the household?

(1) Never
(2) Rarely
(3) Sometimes
(4) Usually
(5) Always

W12 – Overall, how has having [S.C.] in your life affected your family?

(1) Very Positively
(2) Somewhat Positively
(3) Mixed
(4) Somewhat Negatively
(5) Very Negatively

W13 – So far, how has having [S.C.] in your life compared with what you thought it would be like?

(1) Better than you ever expected
(2) About what you expected
(3) More difficult than you ever expected

W14 – Overall, how do you think [S.C.] feels about being adopted?

(1) Feels positive about it

(2) Feels mostly positive about it
(3) Feels neither positive nor negative about it
(4) Feels mostly negative about it
(5) Feels negative about it

W15 – If you [and your spouse/partner] knew everything about [S.C.] before the adoption that you now know, how might that have affected your decision to accept [him/her] for adoption? Would you have…?
(1) Would have definitely accepted the child
(2) Would have probably accepted the child
(3) Would have probably not accepted the child
(4) Would have definitely not accepted the child

W16 – Given your [and your spouse's/partner's] experience of adoption with this child, would you recommend adoption to others?
(1) Yes
(2) No
(3) Depends

W17 – Since the adoption was finalized, has [S.C.] ever lived outside of your home for two weeks or longer?
(1) Yes
(2) No

W17AR – How many times? (follow-up to item W17)

W17B – Was [S.C.]'s time away from home related to problems or conflicts among family members?
(1) Yes
(2) No

W18 – Have you ever thought about ending this adoption?
(1) Yes
(2) No

W19 through **W23HA** excluded from this analysis due to lack of data.

Appendix B – Participant Demographics

Table 1 - base dataset		TYPE OF ADOPTION			Total
		International	Foster	Private	
Derived - Sex of selected child	Male	181	378	399	958
	Female	364	385	382	1131
Total		545	763	781	2089

Table 2 - base dataset		TYPE OF ADOPTION			Total
		International	Foster	Private	
Derived - Sex of selected child	Male	9%	18%	19%	46%
	Female	17%	18%	18%	54%
Total		26%	37%	37%	100%

Identified sex of adoptive parent by adoption type & Collection group

Appendix C – Item by Model

Items by dimension within each Model

	Unidimensional	Multidimensional nested	Multidimensional 3-dimensional	Park, Barth, & Harrington
Dim A	W1, W1A, W2R, W3, W4R, W6R, W7, W8, W12, W13, W14, W15	W1A, W4R, W6R, W12, W14, &W15	W1A, W4R, W6R, W12, W14, &W15	W1A, W2R, W3
Dim B		W1, W2R, W3, W7, W8, & W13	W1, W2R, W3, W7, W8, & W13	W7, W12, W13, W15, W16
Dim C			W5, W9, W16, W17, W17AR*, & W17B**	

Note. List of items in each of the dimensions for the examined models. * Excluded from ConQuest Dimension C analysis. **Excluded from ConQuest Dimension C and 3-Dimensional model analysis.

Appendix D – ConQuest Item Fit Indices

Item Fit Information for Unidimensional ConQuest

	Infit (Weighted)		Outfit (Unweighted)	
Item	MS	T	MS	T
W1A	0.8	-2.4	0.7	-6.3
W4R	0.84	-2.5	0.84	-2.3
W6R	0.86	-2.7	0.87	-2.8
W12	0.78	-3.1	0.73	-6.9
W14	0.91	-2.1	0.91	-2.0
W15	0.87	-0.4	0.34	-20.5
W1	0.86	-0.6	0.45	-16.2
W2R	0.85	-1.7	0.78	-5.4
W3	0.84	-1.1	0.54	-12.6
W7	0.90	-1.9	0.81	-4.7
W8	0.85	-2.0	0.79	-4.3
W13	0.96	-0.6	0.93	-1.7

Note. MS represents mean square and T represents item fit.

Item Fit Information for Dimension A ConQuest

	Infit (Weighted)		Outfit (Unweighted)	
Item	MS	T	MS	T
W1A	0.86	-1.6	0.72	-5.7
W4R	0.90	-1.5	0.89	-1.5
W6R	0.92	-1.5	0.90	-2.0
W12	0.91	-1.4	0.82	-4.4
W14	0.94	-1.3	0.93	-1.3
W15	1.03	0.2	0.99	-0.2

Note. MS represents mean square and T represents item fit.

Item Fit Information for Dimension B ConQuest

	Infit (Weighted)		Outfit (Unweighted)	
Item	MS	T	MS	T
W1	0.92	-1.1	0.73	-6.9
W2R	1.19	3.3	1.35	7.2
W3	0.85	-2.2	0.69	-7.9
W7	0.96	-1.0	0.92	-1.8
W8	0.97	-0.6	0.91	-1.8
W13	0.98	-0.2	1.08	1.7

Note. MS represents mean square and T represents item fit.

Item Fit Information for Dimension C ConQuest

	Infit (Weighted)		Outfit (Unweighted)	
Item	MS	T	MS	T
W5	1.03	0.5	1.03	0.6
W9R	1.04	0.6	1.04	0.6
W16	1.04	0.3	1.36	7.4
W17	1.03	0.3	1.17	3.7
W17AR	1.15	0.7	1.23	1.3

Note. MS represents mean square and T represents item fit.

Item Fit Information for 3-dimensional Model ConQuest

	Infit (Weighted)		Outfit (Unweighted)	
Item	MS	T	MS	T
W1A	0.85	-1.9	0.78	-4.4
W4R	0.92	-1.3	0.91	-1.2
W6R	0.91	-1.7	0.92	-1.7
W12	0.87	-2.1	0.86	-3.4
W14	0.94	-1.3	0.94	-1.3
W15	1.03	0.2	1.01	0.3
W1	1.0	0.1	0.99	-0.3

W2R	1.01	0.2	1.03	0.8
W3	0.99	0	0.86	-3.3
W7	1.02	0.3	1.08	1.8
W8	1.01	0.1	1.05	0.9
W13	1.04	0.6	1.10	2.3
W5	1.02	0.3	1.05	1.0
W9R	0.98	-0.2	0.99	-0.2
W16	0.99	0	1.06	1.3
W17	0.98	0.3	1.0	-0.1

Note. MS represents mean square and T represents item fit.

Item Fit Information for 2-dimensional Model ConQuest

	Infit (Weighted)		Outfit (Unweighted)	
Item	MS	T	MS	T
W1A	0.96	-.05	0.81	-3.9
W4R	1.20	2.6	1.23	3.0
W6R	1.17	3.1	1.16	3.1
W12	1.0	0.1	1.12	2.6
W14	1.16	3.2	1.20	3.9
W15	0.95	-0.2	0.48	-14.8
W1	1.04	0.3	0.58	-11.3

W2R	1.18	2.0	1.01	0.2
W3	0.99	0	0.46	-15.6
W7	0.85	-2.9	0.69	-8.1
W8	0.87	-1.9	0.74	-5.3
W13	0.98	-0.4	0.93	-1.6

Note. MS represents mean square and T represents item fit.

Appendix E – Person Logit Correlations

Person Logit Correlations by Dimension

Model	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Unidimensional Winsteps	1												
2. Unidimensional ConQuest	.76	1											
3. Dim. A Winsteps	.80	.80	1										
4. Dim. A ConQuest	.76	.90	.89	1									
5. Dim. B Winsteps	.93	.71	.65	.63	1								
6. Dim. B ConQuest	.89	.73	.63	.63	.95	1							
7. Dim. C Winsteps	.35	.32	.34	.32	.36	.33	1						
8. Dim. C ConQuest	.35	.33	.33	.33	.35	.34	.92	1					
9. 3-Dim 1st dimensional	-.08*	-.02**	-.23	-.18	.03**	.06**	-.54	-.53	1				
10. 3-Dim 2nd dimensional	-.07*	-.10	-.01**	-.04**	-.08	-.11	.58	.58	-.91	1			
11. 3-Dim 3rdt dimensional	.65	.63	.80	.77	.46	.45	.11	.11	-.18	-.18	1		
12. 2-Dim 1st dimensional	.83	.95	.84	.89	.76	.77	.33	.33	-.03**	-.11	.66	1	
13. 2-Dim 2nd dimensional	.79	.87	.86	.90	.65	.65	.32	.32	-.21	.01**	.72	.91	1

Note. Correlations significant at $p \leq 0.01$ level, except when noted * $p \leq 0.05$ and ** not significant.

Appendix F – Item Difficulty

Item Fit Difficulty by Model

Item	Unidimen. Winsteps	Unidimen. ConQuest	Consec. Winsteps	Consec. ConQuest	3-Dimen. Model	2-Dimen. Model
W1A	1.36	2.09	1.0	1.96	0.7	1.92
W4R	-0.29	0.89	-0.88	0.87	0.3	0.97
W6R	-0.39	0.89	-0.89	0.88	0.33	1.02
W12	0.79	1.55	0.45	1.55	0.57	1.64
W14	-0.5	1.14	-1.01	1.13	0.39	0.97
W15	1.6	2.87	1.34	2.83	2.8	4.31
W1	1.36	2.75	2.61	4.2	2.77	4.13
W2R	0.23	1.62	0.89	3.02	1.65	2.93
W3	0.79	2.25	1.63	3.63	2.28	3.54
W7	-1.73	1.01	-1.83	0.92	1.04	1.49
W8	-1.83	2.5	-1.93	2.97	1.04	4.02
W13	-1.39	1.26	-1.37	2.2	1.37	1.32

W5	-	-	1.21	0.67	0.92	-
W9R	-	-	0.4	0.26	0.44	-
W16	-	-	4.66	2.41	4.09	-
W17	-	-	-5.5	-2.89	4.47	-
W17 AR	-	-	-0.96	0.69	-	-
W17 B	-	-	0.2	-	-	-

Note. Consecutive models combine for this table.

Appendix G – Glossary

(Bond & Fox, 2007)

Ability estimate – The location of a person on a variable, inferred by using the collected observations.

Bias – The difference between the expected value of a sample statistic and the population parameter statistic estimates. It also infers the effects of any factor that the researcher did not expect to influence the dependent variable. (see DIF.)

Calibration – The procedure of estimating person ability or item difficulty by converting raw scores to logits on an objective measurement scale.

Classical test theory – See True score model/Traditional test theory. True score theory is classical in the sense that it is traditional.

Concurrent validity – The validity of a measure determined by how well it performs with some other measure the researcher believes to be valid.

Construct – A single latent trait, characteristic, attribute, or dimension assumed to be underlying a set of items.

Construct validity – Theoretical argument that the items are actual instantiations or operationalizations of the theoretical construct or latent trait under investigation; that is, that the instrument measures exactly what it claims to measure.

Counts – The simple attributions of numerals to record observation. In the Rasch model, raw scores are regarded as counts.

Deterministic – Characteristic of a model that implies the exact prediction of an outcome. Deterministic models explicate the relation between the observed responses and person ability as a causal pattern; for example, Guttman scaling is deterministic – the total score predicts exactly which items were correctly answered. (Cf. Probabilistic.)

Dichotomous – Dichotomous data have only two values such as right/wrong, pass/fail, yes/no, mastery/fail, satisfactory/unsatisfactory, agree/disagree, male/female.

Differential item functioning (DIF) – The loss of invariance of items estimates across testing occasions. DIF is prima facie evidence of items bias.

Error – The difference between an observation and a predication or estimation; the deviation score.

Error estimate – The difference between the observed and the expected response associate with item difficulty or person ability.

Estimation – The Rasch process of using the obtained raw scores to calculate the probable values of person and item parameters.

Fit – The degree of match between the pattern of observed responses and the modeled expectations. The can express either the pattern of responses observed for a candidate on each item (person) or the pattern for each item on all persons (item fit).

Fit statistics – Indices that estimate the extent to which responses show adherence to the modeled expectations.

Fundamental measurement – Physicist Norman Cambel showed that the physical scientists mean by measurement requires an ordering system and the kind of additivity illustrated by physical concatenation. He called this "fundamental measurement" The Rasch model is a special case of additive conjoint measurement, a form of fundamental measurement.

Infit mean square – One of the two alternative measures that indicate the degree of fit of an item or a person (the other being standardized infit). Infit mean square is a transformation of the residuals, the difference between the predicted and the observed, for the residuals, the difference between the predicted and the observed, for easy interpretation. Its expected value is 1. As a rule of thumb, values between 0.70 and 1.30 are generally regarded as acceptable. Values greater than 1.30 are termed mis-fitting, and those less than 0.70 as over fitting.

Infit statistics – Statistics indicating the degree of fit of observations to the Rasch-modeled expectations, weighted to give more value to on-target observations. Infit statistics are more sensitive to irregular inlying patterns and are usually expressed in tow forms: unstandardized as mean square and standardized as *t*.

Infit *t* – One of the two alternative measures that indicate the degree of fit of an item or a person to the Rasch model (the other being infit mean square). The infit *t* (also called standardized infit) is the standardization of fit values to a distribution with a mean of 0 and variance of 1. Values in the range of -2 to +2 are usually held as acceptable ($p < .05$). Values greater than +2 are regarded as mis-fitting, and those less than -2 as overfitting.

Invariance – The maintenance of the identity of a variable from one occasion to the next. For example, item estimates remain stable across suitable sample; person estimate remain stable across suitable test.

Item characteristic curve (ICC) – An ogive-shaped plot of the probabilities of a correct response on an item for any value of the underlying trait in a respondent.

Item difficulty – An estimate of an item's underlying difficulty calculated from the total number of persons in an appropriate sample who succeeded on that item.

Item fit statistics – Indices that show the extent to which each item performance matches the Rasch-modeled expectations. Fitting items imply a unidimensional variable.

Item measure – The Rasch estimate of item difficulty in logits.

Item reliability index – The estimate of the replicability of item placement within a hierarchy of items along the measured variable if these same items were to be given to another sample of comparable ability. Analogous to Cronbach's alpha, it is bounded by 0 and 1.

Item Response Theory (IRT) – A relatively recent development in psychometric theory that overcomes deficiencies of the classical test theory with a family of models to assess model-data fit and evaluate educational and psychological tests. The central postulate of IRT is that the probability of a person's expected response to an item is the joint function of that person's ability, or location on the latent trait, that one or more parameters characterizing the item. The response probability is displayed in the form of an item characteristic curve as a function of the latent trait.

Item separation index – An estimate of the spread or separation of items on the measured variable. It is expressed in standard error units, that is, the adjusted item standard deviation divided by the average measurement error.

Iteration – A repetition. In Rasch analysis computation, the item estimation/person estimation cycle is repeated until a specified condition (convergence criterion) is met.

Latent trait – A characteristic or attribute of a person that can be interred from the observation of the person's behaviors. These observable behaviors display more or less the characteristic, but none of the observation covers all of the trait.

Latent trait theory – See Item response theory.

Likert scale – A widely used questionnaire format in human science research, especially in the investigation of attitudes. Respondents are given statement or prompts and asked to endorse a response from the range of ordered response options, such as "strongly agree," "agree," "neutral," "disagree," or "strongly disagree." (after R. Likert)

Local independence – The items of a test are statistically independent of each subpopulation of examinees whose members are homogeneous with respect to the latent trait measured.

Logit – The unit of measurement that results when the Rasch model is used to transform raw scores obtained from ordinal data to log odds ratios on a common interval scale. The value of 0.0 logits is routinely allocated to the mean of the item difficulty estimates.

Measurement – The location of objects along a single dimension on the basis of observation which add together.

Measurement error – Inaccuracy resulting from a flaw in a measuring instrument – as contracted with other sours of error or unexplained variance.

Measurement precision – The accuracy of any measurement.

Missing data – One or more values that are not available for a subject or case about whom other values are available, for example, a question in a survey that a subject does not answer. The Rasch model is robust in the face of missing data.

Model – A mathematical model is required to obtain measurements from discrete observations.

Muted – Items or person with infit mean square values less than 0.70 or infit *t* values less than -2 are considered muted or overfitting. This indicates less variability in the data than the Rasch model predicts and generally reflects dependency in the data.

95% confidence band – In test or person equating, the interval within the control lines set by the investigator (at $p << .05$) requiring that 95% of measured items or persons should fit the model.

Noisy – Items or persons with infit mean square values greater than 1.30 or infit *t* values greater than +2 are considered noisy or mis-fitting. This indicates more erratic or haphazard performance than the Rasch model predicts.

Nominal scale – A scale in which numerals are allocated to category values that are not ordered. Although this is necessary for measurement, it is not sufficient for any form of scientific measurements. (Cf. Stevens.)

One-parameter item response model (1PL-IRT) – This description of Rasch model highlights the Rasch focus on just one item parameter – difficult- along with the model's membership in the IRT family of data-fitting models. Such a description usually ignores the Rasch model focus on fundamental measurement.

Order – The transitive relationship between values A, B, C, etc., of a variable such that A > B, B > C, A > C, etc.

Order effect – When subjects receive more than one treatment or intervention, the order in which they receive those treatments or interventions might affect the result. To avoid this problem, researchers often use counterbalanced design.

Ordinal scale – A method of comparisons that ranks observations (put them in an order) on some variable and allocated increasing values (e.g., numerals 1, 2, 3 or letters a, b, c, etc.) to that order. The size difference between ranks is not specified. Although this is necessary for measurement, it is not sufficient for any form of scientific measurement. (Cf. Stevens.)

Outfit statistics – Unweighted estimates of degree of fit of responses. These unweighted values tend to be influenced by the off-target observations and are expressed in two forms: unstandardized mean squares and standardized t values.

Overfit – See Muted.

Partial credit analysis – A Rasch model for polytomous data, developed in the work of Geoff Masters (esp.), which allows the number of ordered item categories and/or their threshold values to vary from item to item.

Perfect score – The maximum possible score a respondent can achieve on a given test by answering all of items correctly or endorsing the highest level response category for every item.

Person ability – See Person measure.

Person fit statistics – Indices that estimate the extent to which the responses of any person conform to the Rasch model expectation.

Person measure – An estimate of a person's underlying ability based on that person's performance on a set of items that measure a single trait. It is calculated from the total number of items to which the person responded successfully in an appropriate test.

Person reliability index – The estimate of the replicability of person placement that can be expected if this sample of persons were to be given another set of items measuring the same construct. Analogous to Cronbach's alpha, it is bounded by 0 and 1.

Person separation index – An estimate of the spread or separation of persons on the measured variable. It is expressed in standard error units, that is, the adjusted person standard deviation divided by the average measurement error.

Probabilistic – Given that all the possible influences on person performance cannot be known, the outcomes of the Rasch model are expressed mathematically as probabilities; for example, Rasch measurement is probabilistic – the total score predicts with varying degrees of certainty which items were correctly answered. (See Stochastic.)

Rating scale analysis – A version of the Rasch model, developed in the work of David Andrich (esp.), now routinely used for the sort of polytomous data generated by Likert scales. It requires that every item in a test have the same number of response options, and applies the one set of threshold values to all items on the test.

Raw scores – Scores or counts in their original state that have not been statistically manipulated.

Residual – The residual values represent the difference between the Rasch model's theoretical expectations and the actual performances.

Segmentation – When tests with items at different developmental levels are submitted to Rasch analysis, items representing different stages should be contained to different segments of the scale with a nonzero distance between segments. The items should be mapped in the order predicted by the theory.

Specific Objectivity – The measurement of any person's trait is independent of the dispersion of the set of items used to measure that trait and, conversely, item calibration is independent of the distribution of the ability in the sample of persons who take the test.

Standardized infit – See Infit *t*.

Standardized outfit – Unweighted estimates of the degree of fit of responses. The outfit statistic is routinely reported in its unstandardized (mean square) and standardized (*t* statistics) forms. The acceptable values for *t* range from -2 to +2 ($p < .05$). Values greater than +2 are termed mis-fitting and those less than -2 as overfitting. Compared with the infit statistics, which give more weight to on-target performances, the outfit t statistic is more sensitive to the influence of outlying scores.

Step – See Threshold.

Stochastic – Characteristic of a model that expresses the probabilistic expectations of item and person performance on the construct held to underlie the observed behaviors.

Targeted – The items on the testing instrument match the range of the test candidates' proficiency.

Three-parameter item response model (3PL-IRT) – An item response model that estimates three item parameters – item difficult, item discrimination, and guessing – to better fit the model to the empirical data.

Threshold – The level at which the likelihood of failure to agree with or endorse a given response category (below the threshold) turns to the likelihood of agreeing with or endorsing the category (above the threshold).

Traditional test theory – See True score model/Classical test theory.

True Score model – The model indicates that any observed test score could be envisioned as the composite of two hypothetical components: a true score and a random error component.

Two-parameter item response model (2PL-IRT) – An item response model that estimates two item parameters – item difficulty and item discrimination – to better fit the model to the empirical data.

Unidimensionality – A basic concept in scientific measurement that one attribute of an object (e.g., length, width, weight, temperature, etc.) be measured at a time. The Rasch model requires a single construct to be underlying the items that from hierarchical continuum.

Validity – Evidence gathered to support the inference made from responses to explicate the meaningfulness of a measured construct through examining person fit, item fit, and item and person ordering.

Variable – An attribute of the object of study that can have a variety of magnitudes. The operationalization of a scale to measure these values is termed variable construction. A variable is necessarily unidimensional.

Appendix H – Institutional Review Board Approval

UNIVERSITY of DENVER
RESEARCH & SPONSORED PROGRAMS

DATE:	May 17, 2017
TO:	Michael Furno
FROM:	University of Denver (DU) IRB
PROJECT TITLE:	[1066213-1] Rasch and Multidimensional Rasch Analysis of Relational Well-Being within the National Survey of Adoptive Parents of 2007
SUBMISSION TYPE:	New Project - Exempt from IRB Review
ACTION:	**EXEMPTION GRANTED**
DECISION DATE:	May 17, 2017
EXEMPTION VALID THROUGH:	May 16, 2020
RISK LEVEL:	Minimal Risk
REVIEW CATEGORY:	Exemption category # *4*
	Exemption Category 4: Collection or Study of Existing Data - Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects.

Thank you for your submission of Exemption Request materials for this project. The University of Denver IRB has determined this project is **EXEMPT FROM IRB REVIEW** according to federal regulations. This exemption was granted based on appropriate criteria for granting an exemption and a study design wherein the risks have been minimized.

Exempt status means that the study does not vary significantly from the description that has been provided and further review in the form of filing an annual Continuing Review/Progress Report is not required.

Please note that maintaining exempt status requires that (a) risks of the study remain minimal; (b) that anonymity or confidentiality of participants, or protection of participants against any increased risk due to the internal knowledge or disclosure of identity by the researcher, is maintained as described in the application; (c) that no deception is introduced, such as reducing the accuracy or specificity of information about the research protocol that is given to prospective participants; (d) the research purpose, sponsor, and recruited study population remain as described; and (e) the principal investigator (PI) continues and is not replaced.

If changes occur in any of the features of the study as described, this may affect one or more of the conditions of exemption and may warrant a reclassification of the research protocol from exempt and require additional IRB review. For the duration of your research study, any changes in the proposed study must be reviewed by the University of Denver IRB before implementation of those changes.

This exemption has been granted for a three-year time period. The DU Human Research Protection Program (HRPP) will retain a copy of this correspondence within our records and will administratively