Applying the Nominal Response Model Within a Longitudinal Framework to Construct the Positive Family Relationships Scale

Educational and Psychological
Measurement
2015, Vol. 75(6) 901–930
© The Author(s) 2015
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0013164414568717
epm.sagepub.com

Kathleen Suzanne Johnson Preston¹, Skye N. Parral¹,
Allen W. Gottfried¹, Pamela H. Oliver¹,
Adele Eskeles Gottfried², Sirena M. Ibrahim¹, and
Danielle Delany¹

Abstract

A psychometric analysis was conducted using the nominal response model under the item response theory framework to construct the Positive Family Relationships scale. Using data from the Fullerton Longitudinal Study, this scale was constructed within a long-term longitudinal framework spanning middle childhood through adolescence. Items tapping this construct were completed annually by mothers when children were aged 9 to 17 years. The scale measures a construct customized for each age with uniquely informative items consisting exclusively of discriminating response categories as evaluated using the nominal response model. By using longitudinal data, the study is innovative in implementing the method of test equating to a psychological construct across nine assessments spanning 8 years. Future research pertaining to validity is discussed.

Keywords

nominal response model, item response theory, Positive Family Relationships scale, longitudinal research, scale construction

¹California State University, Fullerton, CA, USA

²California State University, Northridge, CA, USA

Corresponding Author:

Kathleen Suzanne Johnson Preston, Department of Psychology, California State University, Fullerton, PO Box 6846, Fullerton, CA 92834-6846, USA.

Email: kpreston@fullerton.edu

Psychological scales are traditionally developed under a classical test theory framework using cross-sectional data (Reise, Ainsworth, & Haviland, 2005). Although, item response theory (IRT) has been used as the primary method in educational scale development and assessment (Harvey & Hammer, 1999), the flexibility of IRT is increasingly being more broadly recognized, thus bringing these analytic methods to the forefront of scale development and evaluation for measurement of psychological constructs (Reise & Henson, 2003). In fact, published IRT-based articles found within the PsycINFO database have increased approximately 25% over the past 10 years, with 658 articles during 2004-2006 to 866 articles during 2012-2014.

In contrast to the ubiquitously used cross-sectional approach to scale construction, in the present research, a long-term longitudinal measure of the construct under investigation was developed by employing IRT. Unique to the present research, the customized scale was constructed over an extensive time frame. By using longitudinal rather than cross-sectional data for scale construction, unnecessary between-subject variability is eliminated, which results in a scale measuring the latent trait more precisely through reduction of error variance. Furthermore, using longitudinal data for scale construction eliminates differences across measurement years due to possible cohort effects.

In measuring a psychological construct over time, there is no assumption that the same items should reflect the construct at all ages (see, e.g., Sy, Gottfried, & Gottfried, 2013). That is, whereas the underlying construct remains time invariant, items used to measure the construct may become more or less informative as individuals advance in age (e.g., from childhood through adolescence). For instance, measuring an attribute of families across children's development may require different items at different ages. Therefore, an item suitable to assess family functioning with young children could be differentially informative than for families with adolescents. Constructing a scale that is customized for specific ages furnishes the most accurate measure and best representation of the latent construct. Conventionally, it has been necessary for individuals to respond to the same set of items in order to evaluate change over time. However, a benefit of IRT is the ability to link individuals' scores over time when they have responded to scales containing at least one common item (Kelderman, 1988), thereby providing the opportunity to assess both individual and normative change based on scales containing exclusively the most informative items across ages.

Psychological scales are commonly assigned 5-, 7-, or even 9-point Likert-type response formats without any justification for the number of response options. It has been assumed that the response options are ordered and function equivalently across the entire range of the construct, but explicit evaluation of the response options is lacking. There are various models under the IRT framework to evaluate polytomous response formats such as the graded response model (GRM; Samejima, 1969), the multiple-choice model (MCM; Thissen, Steinberg, and Fitzpatrick, 1989), the generalized partial credit model (GPCM; Muraki, 1992), the rating scale model (RSM; Andrich, 1978), and the nominal response model (NRM; Bock, 1972, 1997), with the

three latter models being from the divide-by-total family of models, all of which have contributed in their own right to scale construction. In this article, the latter model has been selected because of its unique ability to empirically evaluate the functioning of each within-item response option in addition to the overall functioning of the items (Embretson & Reise, 2000). As argued by Reise (Reise, 2009; Reise & Waller, 2009) some traits may not be fully continuous and consequently respondents cannot make distinctions at all levels of a continuum.

Preston, Reise, Cai, and Hays (2011) demonstrated the NRM, the most general of the divide-by-total family of polytomous IRT models, as a useful method in identifying differentially and nonfunctioning response options without imposing parameter restrictions, as is done in the more constrained divide-by-total models (e.g., GPCM) nested within the NRM (see Preston & Reise, 2014; Ostini & Nering, 2006). Furthermore, Preston and Reise (2014) illustrate applications of the NRM in diagnosing category functioning in ordered polytomous data. The NRM uniquely provides category-level information, which allows for the empirical evaluation of response format functioning by assessing whether an ordered item contains unnecessary response options (Preston & Reise, 2014). Identifying nonfunctioning categories via NRM suggests that the response format should be modified to include only informative categories. The purpose of this research was to create a scale that (a) measured the same latent construct across multiple assessment waves, (b) used long-term longitudinal data, (c) was customized for each age so as to contain only informative items, and (d) had an empirically derived response format as estimated under the NRM.

Positive Family Relationships

Psychologists from various disciplines have identified the family as a key factor in multiple domains of adjustment across the life span. Historically, the overwhelming body of research has focused on the negative behavioral aspects of the family pertaining to adverse developmental outcomes. However, in recent years, researchers began to broaden the focus to beneficial family factors related to healthy adjustment and well-being of youth (Benson, Scales, Hamilton, & Sesma, 2006). Positive family functioning during childhood and adolescence has long-term associations with beneficial individual functioning into adulthood (Paradis et al., 2011). In this study, the NRM is applied within a long-term longitudinal framework encompassing nine annual assessment waves traversing an 8-year interval to construct a positive family relationships (PFR) scale. To the knowledge of the present researchers, the current study is the first application of this approach using IRT to develop a scale measuring PFR.

Method and Results

Participants

The present study used data from the Fullerton Longitudinal Study, an ongoing investigation of children from infancy through early adulthood, ages 1 through 29

years (A. E. Gottfried, Marcoulides, Gottfried, & Oliver, 2013; A. W. Gottfried & Gottfried, 1984; Guerin, Gottfried, Oliver, & Thomas, 2003). This investigation was initiated with 130 infants and their families. The infants were selected from birth notifications of hospitals surrounding the university. All children who entered the study had been full-term infants of normal birth weight and had no neurological or visual abnormalities. The current analyses used data collected from the mothers when children were ages 9 to 17 years. Socioeconomic status of the families was assessed with the Hollingshead Four-Factor Index of Social Status (A. W. Gottfried, Gottfried, Bathurst, Guerin, & Parramore, 2003; Hollingshead, 1975). This extensively used index is based on parents' educational attainment and occupational status. Socioeconomic status of the families varied, ranging from semiskilled workers with no high school degree through professionals. The mean Hollingshead Social Status Index was 45.6 ($SD = 11.9$) at the beginning of the study and 48.6 ($SD = 11.4$) at the 17-year assessment. Participants included European American (90%) as well as other ethnic backgrounds (10%). This represented the demographics of the region at the onset of the study. The percentages of males and females were 52 and 48, respectively.

Throughout the course of investigation, the retention rate of participation was substantial with at least 80% of the study sample returning at any assessment wave; and, of those children assessed at age 9 years, at least 93% furnished data at any given assessment wave through 17 years. As for mothers, the return rate was at least 86% across this time frame. Thus, missing data were minimal and there was no evidence of attrition bias in the course of investigation (Guerin et al., 2003).

Positive Family Relationships Scale Development

Positive Family Relationships. Herein we define PFR as family members getting along well and supporting each other. An initial pool of 25 items assessing support, agreement, helpfulness, unity, and discord, was created to measure positive family relationships. Items were constructed with responses along a 6-point scale from 1 (*never*) to 6 (*always*). Scale instructions read, "For each of the following statements, circle the number which describes your family most accurately". These items were administered to the mothers annually from ages 9 through 17 years. Mothers were selected for three reasons: (a) they were the parent who most often brought the children to the university laboratory for the assessments, (b) the children resided predominately or at least halftime with their mothers in situations of separation or divorce, and (c) mothers were most available and had the greatest fund of knowledge, thereby providing a more consistent perspective of the family relationships.

Nominal Response Model Estimation. All items on the PFR scale were scored so that higher values indicated more positive ratings of family relationships. Preliminary exploratory data analysis (e.g., category response frequencies) revealed generally a small proportion of individuals who responded in the extreme categories (i.e. never, and always). For proper estimation of the NRM, responses were recoded to contain

at least 10% of the individuals per category without altering the integrity of the endorsed category, as recommended by Preston et al. (2011). For instance, if an item scored as 1, 2, 3, 4, 5, 6, contained the percentages for each category of 0%, 3%, 9%, 22%, 63%, and 3%, respectively, the first three categories would be combined into the lowest response option and the last two categories would be combined together for the highest response option. This would result in a three-category item (i.e., response options coded as 1, 1, 1, 2, 3, 3). The consolidation process rendered no problem in determining the appropriate direction as there was no neutral option with the six response categories. Estimation of the initial pool of 25 items was conducted on resulting data with responses recoded to contain at least 10% of individuals responding in each response category.

The PFR scale items were modeled with the NRM using flexMIRT (Cai, 2012). All IRT analyses were conducted with program defaults for maximum likelihood estimation, specifying the estimation of the Fisher information function using 81 quadrature points for stability of the standard errors. The supplemented EM (expectation maximization) algorithm was used to improve accuracy in estimation with small sample sizes (Cai, 2008) and expected a posteriori θ score estimation, within FlexMIRT (Cai, 2012).

Category Boundary Discriminations and Intersections. The NRM estimates category boundary discrimination (CBD) parameters, which determine the amount of relative information provided by a response in a particular category versus a response in an adjacent category. The size of a CBD parameter indicates the amount of relative information provided by adjacent response categories (e.g., the degree to which a response in Category 3 vs. Category 2 differentiates among people on the latent trait). Large positive CBD parameters indicate that the distinction between two adjacent categories is highly informative, whereas a near-zero value indicates that the distinction between the categories is meaningless because individuals cannot differentiate between the response options.

Also estimated under the NRM are intersection parameters, which identify the point on the latent trait where an individual is likely to respond in adjacent categories. A useful way to visualize the functioning of an item and its individual categories is to use the CBDs and intersections to produce plots of the category response curves (CRCs). CRCs represent the relationship between an individual's latent trait level, presented on the horizontal axis, and that individual's conditional probability of responding in a particular category, presented on the vertical axis (Preston et al., 2011).

Category Response Curves and Information. An example set of CRCs and item information functions (IIFs) is displayed in Figures 1 and 2, respectively, for an item with CBD parameters of 1.5, 1.5, and 0.5, and intersections of -1.5 , 0 , and 1.5 . Vertical dashed lines have been superimposed on the figure to denote the location of the intersections of each CRC, which indicate the location on the latent trait where curves

Figure 1. Category response curves for an item modeled under the nominal response model with varying category boundary discrimination parameters of $a_1^* = 1.5$, $a_2^* = 1.5$, $a_3^* = 0.5$.

representing adjacent categories meet and a particular category become less likely and the adjacent category become most likely to be endorsed. For instance, curves for Categories 1 and 2 intersect at 1.5 standard deviations below the mean, which indicates that individuals below 1.5 standard deviations below the mean on the construct are most likely to endorse Category 1, individuals at 1.5 standard deviations below the mean are equally likely to endorse Categories 1 and 2, and individuals above 1.5 standard deviations below the mean are most likely to endorse Category 2 until the mean, where the curves for Categories 2 and 3 intersect. For model identification purposes, the probability functions for the lowest (Category 1) and highest (Category 4) response categories are required to be monotonically decreasing and increasing, respectively, with the inflection point of the curve indicating the point on the latent trait at which the category is most likely to be endorsed, and, thereby, most informative. The remaining categories' probability functions are a parabolic shape with the peak of the curve occurring at the midpoint between the locations of the category intersections. The steepness of the curves at each intersection is determined

Figure 2. Item and category information for an item modeled under the nominal response model with varying category boundary discrimination parameters of $a_1^* = 1.5$, $a_2^* = 1.5$, $a_3^* = 0.5$.

by the CBD parameter. As can be seen, the distinctions between Categories 1 and 2, and Categories 2 and 3 function equivalently with CBDs of 1.5. However, the distinction between Categories 3 and 4 is less informative, with a CBD of 0.5, visually producing curves that intersect at a shallower angle.

Figure 2 displays category and item information for the same item. The item information is represented by the overarching curve and is merely the sum of the category information curves plotted underneath. The CBDs determine the steepness of each category information function, with the peak of the function occurring at the location of the intersection. Specifically, as the CBD gets smaller for the distinction between Categories 3 and 4, the relevant category distinction provides less information because the distinction between two categories is less meaningful. In such cases, a researcher may want to consider revising the item response format or rethinking the content used to denote the response categories. As can be seen, the amount of information an item provides is contingent on the CBDs, and the spread of the

information across the latent trait is determined by the location of the intersections. Furthermore, test information is produced by summing all IIF. Therefore, constructing a scale with optimal functioning, at the category level, is crucial to maximizing the ultimate information the scale provides. CRCs, IIFs, and test information functions (TIF) were produced for the initial pool of items at each age using R (R Core Team, 2014).

Tables 1 to 5 contain the final version of the PFR scale, including item content, CBD parameters and intersection parameters, and standard errors for each age. The parameters shown in the table represent the response option modification conducted for each item at each age. Although the sample size used for parameter estimation under the NRM can be considered small, as can be seen in Tables 1 to 5, the standard errors produced via the supplemented EM algorithm (Cai, 2008) are not unreasonable in comparison with other standard errors estimated under the NRM (e.g., Preston et al., 2011), indicating the appropriateness of using the NRM for the purposes of constructing the PFR scale. For each age, review of item content and adequacy of item information, as estimated under the NRM, determined inclusion of an item in the scale. Consistent with Ura, Preston, and Mearns (2014), all included items produced relative information peaks at or higher than 0.2; those items with information below 0.2 were eliminated from the scale. For instance, the item, "Family members adhere to specifically defined roles in our household," did not provide information at any age, thus empirically appeared incongruent with the measure of PFR and therefore the item was eliminated. However, some items that were initially uninformative and removed, such as Item 3, "Input from all family members is considered when planning family activities," became more informative items as the children advanced in age, and thus were included in subsequent ages. Last, 10 items consistently provided information across all ages, indicating continuity in the measurement of the PFR construct. In summary, the resulting scales constructed for each age contain only uniquely informative items. Items are uniquely informative because other items with redundant information were evaluated and only items providing the most information across the latent trait were included in the final iteration of the scale at a given age.

Item Inclusion Across Years. Figures 3 and 4 display the CRCs for Item 3 at ages 11 and 15 years, respectively. In the figures (moving from left to right on the horizontal axis), Category 1 is represented by a dashed line, Category 2 by a dotted line, Category 3 by a dashed-dotted line, and Category 4 by a long-dashed line (indicated by red, green, blue, and cyan, respectively as presented in the online version). As discussed previously, high CBD parameter values indicate that the categories are distinct, with these CRC intersections indicating that a clear choice is being made between the adjacent categories and are more informative about the respondent. For instance, the CBDs for age 15 years of 1.00, 0.62, and 0.41 are more discriminating compared with age 11 years with CBD values of 0.38, 0.77, and 0.09; thus, the category functioning for age 15 years is superior to age 11 years, as can be seen in the CRC plots (see Figures 3 and 4). The larger CBD values produced for Item 3 at age

Table 1. Positive Family Relationship Scale Category Boundary Discrimination (CBD) and Intersection (Int) Parameters as Estimated Under the Nominal Response Model for Ages 9 Through 17 Years: Items 1 to 5 (R Indicates Reverse Coded Item).

Item	CBD parameters				Intersections			
	CBD ₁ (SE)	CBD ₂ (SE)	CBD ₃ (SE)	CBD ₄ (SE)	Int ₁ (SE)	Int ₂ (SE)	Int ₃ (SE)	Int ₄ (SE)
Item 1: Members of our family get along well together								
Age 9	1.94 (0.55)	2.12 (0.48)			-0.77 (0.50)	1.55 (0.47)		
Age 10	1.54 (0.46)	3.54 (1.30)			-0.73 (1.08)	1.45 (1.02)		
Age 11	1.40 (0.70)	2.05 (0.83)			-2.01 (0.54)	-0.53 (0.56)		
Age 12	1.00 (0.64)	1.68 (1.28)			-1.49 (0.37)	-0.74 (0.44)		
Age 13	1.81 (0.75)	2.55 (0.65)			-1.43 (0.52)	-0.10 (0.44)		
Age 14	3.04 (1.45)	1.74 (0.29)			-1.42 (0.95)	-0.45 (0.92)		
Age 15	3.27 (0.87)				-0.16 (0.49)			
Age 16	1.62 (0.85)	2.73 (1.03)			-1.17 (0.54)	-0.51 (0.54)		
Age 17	1.72 (0.86)	2.37 (0.80)			-1.37 (0.52)	-0.37 (0.52)		
Item 2: Family members keep problems and feelings to themselves (R)								
Age 9	0.86 (0.37)	0.49 (0.45)			-2.19 (0.23)	3.24 (0.25)		
Item 3: Input from all family members is considered when planning family activities								
Age 13	1.49 (0.51)				-2.01 (0.61)			
Age 14	0.42 (0.28)	1.21 (2.37)			-0.93 (0.22)	1.11 (0.25)		
Age 15	1.00 (0.46)	0.62 (0.48)	0.41 (0.39)		-1.13 (0.17)	-0.55 (0.33)	2.37 (0.18)	
Age 16	0.88 (0.34)	0.46 (0.46)			-0.77 (0.19)	2.02 (0.23)		
Age 17	0.86 (0.32)	0.76 (0.62)			-0.27 (0.24)	2.20 (0.28)		

(continued)

Table 1. (continued)

Item	CBD parameters				Intersections			
	CBD ₁ (SE)	CBD ₂ (SE)	CBD ₃ (SE)	CBD ₄ (SE)	Int ₁ (SE)	Int ₂ (SE)	Int ₃ (SE)	Int ₄ (SE)
Item 4: Rules and regulations are part of our family's routine								
Age 16	0.31 (0.27)	0.78 (2.79)			-0.48 (0.20)	1.59 (0.25)		
Item 5: There is interpersonal conflict in our home (R)								
Age 9	3.00 (0.99)	2.85 (0.48)			-0.17 (0.62)	1.12 (0.56)		
Age 10	3.15 (1.53)	2.38 (0.41)	1.24 (0.29)		-0.89 (0.68)	-0.16 (0.88)	1.96 (0.33)	
Age 11	1.10 (0.47)	2.04 (0.95)	0.79 (0.70)		-1.31 (0.29)	-0.10 (0.50)	2.59 (0.23)	
Age 12	2.15 (0.55)				-0.01 (0.35)			
Age 13	2.24 (0.84)	2.36 (0.50)			-0.56 (0.44)	0.08 (0.35)		
Age 14	1.16 (0.51)	1.37 (0.67)			-0.78 (0.27)	-0.23 (0.30)		
Age 15	3.47 (1.02)				0.28 (0.55)			
Age 16	0.97 (0.49)	1.34 (0.96)	1.63 (1.17)		-0.56 (0.32)	-0.09 (0.54)	1.52 (0.24)	
Age 17	3.27 (1.12)	1.57 (0.62)			0.28 (0.55)	1.46 (0.53)		

Table 2. Positive Family Relationship Scale Category Boundary Discrimination (CBD) and Intersection (Int) Parameters as Estimated Under the Nominal Response Model for Ages 9 Through 17 Years: Items 6 to 10 (R Indicates Reverse Coded Item).

Item	CBD parameters				Intersections			
	CBD ₁ (SE)	CBD ₂ (SE)	CBD ₃ (SE)	CBD ₄ (SE)	Int ₁ (SE)	Int ₂ (SE)	Int ₃ (SE)	Int ₄ (SE)
Item 6: Our family does things together rather than separately								
Age 13	0.85 (0.30)				-1.41 (0.29)			
Age 14	0.81 (0.30)	0.51 (0.52)			-0.70 (0.20)	2.25 (0.23)		
Age 15	0.76 (0.31)	0.88 (0.72)			-0.46 (0.19)	1.09 (0.23)		
Age 16	0.99 (0.35)	0.51 (0.47)			-0.09 (0.21)	2.08 (0.26)		
Age 17	1.74 (0.75)	1.50 (0.43)			-1.44 (0.44)	0.07 (0.43)		
Item 7: Family outings and activities are planned by one person in our family (R)								
Age 10	0.86 (0.33)				1.92 (0.33)			
Age 11	0.66 (0.30)	0.46 (0.76)			0.74 (0.16)	1.07 (0.24)		
Item 8: People in our family criticize each other (R)								
Age 9	1.88 (0.78)	3.34 (0.91)			-0.76 (0.43)	-0.06 (0.39)		
Age 10	1.87 (0.72)	1.70 (0.45)	0.80 (0.35)		-0.99 (0.38)	0.07 (0.52)	2.35 (0.23)	
Age 11	2.19 (0.69)	1.56 (0.34)			-0.78 (0.34)	0.26 (0.33)		
Age 12	1.13 (0.54)	3.98 (2.27)			-1.05 (0.47)	0.42 (0.47)		
Age 13	1.12 (0.66)	1.18 (0.84)	3.42 (2.09)		-1.66 (0.41)	-0.83 (0.63)	0.51 (0.27)	
Age 14	0.60 (0.40)	1.68 (2.13)			-1.12 (0.23)	0.03 (0.29)		
Age 15	1.16 (0.45)	2.22 (0.92)			-0.89 (0.30)	0.54 (0.30)		
Age 16	1.86 (0.73)	1.93 (0.49)			-1.33 (0.41)	0.27 (0.41)		
Age 17	1.43 (0.60)	2.22 (0.84)			-0.26 (0.29)	0.15 (0.33)		

(continued)

Table 2. (continued)

Item	CBD parameters				Intersections			
	CBD ₁ (SE)	CBD ₂ (SE)	CBD ₃ (SE)	CBD ₄ (SE)	Int ₁ (SE)	Int ₂ (SE)	Int ₃ (SE)	Int ₄ (SE)
Item 9: Family members support each other								
Age 9	1.84 (0.60)	0.84 (0.23)			-1.07 (0.30)	0.61 (0.32)		
Age 10	2.13 (0.72)	1.63 (0.33)			-1.32 (0.48)	0.71 (0.41)		
Age 11	1.56 (0.51)	1.55 (0.44)			-1.55 (0.35)	0.63 (0.34)		
Age 12	2.23 (0.81)	0.43 (0.16)			-1.31 (0.42)	2.00 (0.44)		
Age 13	1.38 (0.43)	0.54 (0.28)			-0.78 (0.24)	1.48 (0.25)		
Age 14	1.54 (0.54)	1.66 (0.51)			-1.23 (0.34)	0.75 (0.33)		
Age 15	1.40 (1.07)	2.45 (1.48)	1.24 (0.75)		-1.36 (0.48)	-0.85 (0.78)	1.48 (0.37)	
Age 16	1.62 (0.60)	1.42 (0.44)			-1.04 (0.31)	0.42 (0.32)		
Age 17	1.63 (0.53)	1.48 (0.43)			-0.74 (0.31) ³	0.98 (0.31)		
Item 10: People in our family tend to do their own thing (R)								
Age 11	0.39 (0.38)	0.28 (1.19)	0.55 (1.59)		-3.21 (0.15)	-0.25 (0.29)	0.00 (0.17)	
Age 12	0.60 (0.41)	0.87 (1.16)			-3.38 (0.24)	0.92 (0.27)		
Age 13	1.08 (0.53)	0.79 (0.51)			-1.56 (0.30)	-1.05 (0.34)		
Age 14	0.74 (0.45)	0.65 (0.71)			-0.07 (0.23)	2.34 (0.26)		
Age 16	0.96 (0.32)				0.30 (0.26)			
Age 17	0.76 (0.37)	0.52 (0.61)			-0.58 (0.17)	-0.48 (0.24)		

Table 3. Positive Family Relationship Scale Category Boundary Discrimination (CBD) and Intersection (Int) Parameters as Estimated Under the Nominal Response Model for Ages 9 Through 17 Years: Items 11 to 15 (R Indicates Reverse Coded Item).

Item	CBD parameters				Intersections			
	CBD ₁ (SE)	CBD ₂ (SE)	CBD ₃ (SE)	CBD ₄ (SE)	Int ₁ (SE)	Int ₂ (SE)	Int ₃ (SE)	Int ₄ (SE)
Item 11: Personal issues are openly discussed in our home								
Age 9	0.84 (0.49)	0.40 (0.54)	0.10 (0.33)		-1.42 (0.17)	-1.48 (0.33)	3.40 (0.18)	
Age 11	0.95 (0.36)				1.48 (0.30)			
Age 14	0.65 (0.45)	1.09 (1.16)			-2.55 (0.25)	1.58 (0.27)		
Age 15	0.49 (0.28)	1.13 (1.80)			0.39 (0.23)	1.33 (0.28)		
Age 16	0.78 (0.35)	1.14 (0.93)			-0.64 (0.21)	0.75 (0.25)		
Age 17	0.84 (0.33)	1.17 (0.83)			-0.32 (0.26)	1.50 (0.30)		
Item 12: Family members disagree with each other (R)								
Age 9	0.43 (0.44)	0.63 (1.92)	1.24 (3.11)		-2.49 (0.18)	-0.10 (0.34)	0.90 (0.18)	
Age 10	0.36 (0.39)	0.94 (3.30)	1.13 (3.71)		-3.03 (0.26)	-0.53 (0.40)	1.58 (0.19)	
Age 11	0.83 (0.36)	0.45 (0.51)	0.54 (0.59)		-1.31 (0.17)	-0.04 (0.31)	1.93 (0.17)	
Age 12	0.98 (0.46)	0.77 (0.58)	2.14 (1.47)		-1.50 (0.37)	0.35 (0.34)	1.16 (0.25)	
Age 13	1.43 (0.57)	0.99 (0.39)	1.15 (0.50)		-1.41 (0.29)	0.53 (0.48)	1.46 (0.21)	
Age 14	0.63 (0.41)	0.73 (1.08)	1.06 (1.39)		-1.24 (0.21)	-0.40 (0.38)	1.67 (0.29)	
Age 15	0.39 (0.39)	1.61 (1.96)	1.23 (3.41)		-2.36 (0.21)	-0.08 (0.39)	1.55 (0.18)	
Age 16	1.37 (0.52)	0.79 (0.41)	1.36 (0.60)		-0.95 (0.29)	0.53 (0.47)	1.25 (0.22)	
Age 17	1.59 (0.57)	0.68 (0.43)	0.81 (0.36)		-0.81 (0.23)	0.10 (0.41)	1.83 (0.20)	

(continued)

Table 3. (continued)

Item	CBD parameters				Intersections			
	CBD ₁ (SE)	CBD ₂ (SE)	CBD ₃ (SE)	CBD ₄ (SE)	Int ₁ (SE)	Int ₂ (SE)	Int ₃ (SE)	Int ₄ (SE)
Item 13: We are strict about family members adhering to rules and regulations set up in our family								
Age 10	0.58 (0.24)				0.17 (0.22)			
Age 16	1.11 (0.46)				2.26 (0.49)			
Item 14: When one person is overwhelmed by responsibilities, another family member spontaneously pitches in to help								
Age 9	0.65 (0.41)	0.79 (1.08)	0.55 (0.79)		-0.66 (0.16)	-0.38 (0.30)	1.95 (0.18)	
Age 10	1.04 (0.36)	1.56 (0.85)			0.53 (0.41)	1.27 (0.39)		
Age 11	0.52 (0.35)	1.26 (2.04)	1.40 (2.12)	2.55 (3.59)	-1.92 (0.44)	-0.63 (1.15)	-0.22 (0.42)	1.00 (0.25)
Age 12	1.02 (0.33)				-0.83 (0.85)			
Age 13	1.07 (0.32)				0.64 (0.28)			
Age 14	0.62 (0.45)	1.17 (1.87)	2.44 (3.69)		-0.26 (0.41)	0.22 (0.71)	1.14 (0.29)	
Age 15	1.08 (0.46)	1.83 (0.90)			-1.47 (0.28)	0.42 (0.29)		
Age 16	0.90 (0.42)	0.75 (0.67)	0.99 (0.80)		-1.27 (0.20)	1.17 (0.38)	0.91 (0.21)	
Age 17	1.45 (0.50)	0.70 (0.36)			0.20 (0.20)	-0.14 (0.28)		
Item 15: People in our family have such busy schedules that we rarely know what others are doing (R)								
Age 11	1.29 (0.39)				0.74 (0.30)			
Age 12	0.52 (0.30)				-1.13 (0.17)	0.87 (0.22)		
Age 13	1.18 (0.37)	0.75 (1.20)			-0.42 (0.19)	2.05 (0.24)		
Age 14	0.78 (0.46)	0.29 (0.56)	1.04 (0.94)		-1.38 (0.25)	-0.93 (0.32)	0.57 (0.18)	
Age 15	0.76 (0.32)	1.09 (0.88)			-0.39 (0.20)	0.91 (0.24)		
Age 16	0.71 (0.34)	1.29 (0.90)			0.82 (0.17)	0.53 (0.26)		
Age 17	0.61 (0.34)	0.84 (1.20)			1.11 (0.18)	0.49 (0.27)		

Table 4. Positive Family Relationship Scale Category Boundary Discrimination (CBD) and Intersection (Int) Parameters as Estimated Under the Nominal Response Model for Ages 9 Through 17 Years: Items 16 to 20 (R Indicates Reverse Coded Item).

Item	CBD parameters				Intersections			
	CBD ₁ (SE)	CBD ₂ (SE)	CBD ₃ (SE)	CBD ₄ (SE)	Int ₁ (SE)	Int ₂ (SE)	Int ₃ (SE)	Int ₄ (SE)
Item 16: Daily happenings of family members are topics of conversation in our home								
Age 10	0.77 (0.39)	0.95 (0.82)			-1.91 (0.25)	0.61 (0.25)		
Age 13	0.56 (0.30)	0.41 (0.71)			-1.14 (0.16)	0.68 (0.21)		
Age 14	1.12 (0.42)	0.88 (0.47)			-0.26 (0.22)	0.22 (0.25)		
Age 15	0.99 (0.52)	1.44 (0.88)			-2.08 (0.29)	0.62 (0.31)		
Age 16	0.61 (0.40)	1.32 (1.24)			-2.31 (0.23)	0.61 (0.27)		
Age 17	0.74 (0.37)	0.91 (0.84)			-0.81 (0.19)	0.10 (0.25)		
Item 17: We argue in our family (R)								
Age 9	1.52 (0.54)	1.67 (0.52)			-0.41 (0.27)	0.38 (0.28)		
Age 10	0.99 (0.40)	1.78 (0.98)			-0.12 (0.33)	0.60 (0.29)		
Age 11	0.74 (0.30)	0.92 (0.80)			-0.66 (0.20)	1.22 (0.24)		
Age 12	1.23 (0.49)	2.20 (0.94)			-0.08 (0.30)	0.55 (0.33)		
Age 13	1.84 (0.87)	1.46 (0.51)	1.21 (0.41)		-1.02 (0.57)	-0.44 (0.52)	0.66 (0.24)	
Age 14	1.10 (0.45)	1.85 (0.87)			-1.76 (0.34)	1.06 (0.35)		
Age 15	0.72 (0.55)	0.90 (1.19)	1.33 (1.56)		-2.22 (0.21)	-0.18 (0.39)	0.70 (0.20)	
Age 16	2.65 (0.91)	1.59 (0.28)			-1.17 (0.49)	0.98 (0.47)		
Age 17	1.63 (0.71)	0.96 (0.38)	0.61 (0.28)	-1.32 (0.28)	0.01 (0.41)	0.57 (0.20)		

(continued)

Table 4. (continued)

Item	CBD parameters				Intersections			
	CBD ₁ (SE)	CBD ₂ (SE)	CBD ₃ (SE)	CBD ₄ (SE)	Int ₁ (SE)	Int ₂ (SE)	Int ₃ (SE)	Int ₄ (SE)
Item 19: There is an atmosphere of uneasiness in our home (R)								
Age 9	1.59 (0.56)	1.21 (0.36)			-1.14 (0.29)	0.40 (0.30)		
Age 10	0.59 (0.36)	1.33 (1.66)			-1.69 (0.26)	0.57 (0.24)		
Age 11	2.10 (0.59)	0.66 (0.22)			-0.89 (0.29)	0.88 (0.32)		
Age 12	1.65 (0.56)	1.27 (0.37)			-1.22 (0.30)	1.03 (0.32)		
Age 13	2.20 (1.12)	1.65 (0.75)	1.21 (0.35)		-1.06 (0.38)	-0.68 (0.65)	0.77 (0.30)	
Age 14	1.82 (0.62)	1.64 (0.43)			-1.03 (0.34)	0.66 (0.33)		
Age 15	2.23 (0.78)	1.44 (0.32)			-0.61 (0.32)	0.36 (0.33)		
Age 16	2.27 (0.78)	1.34 (0.29)			-1.01 (0.38)	0.72 (0.38)		
Age 17	1.61 (0.55)	1.89 (0.59)			-0.42 (0.32)	0.70 (0.32)		
Item 20: Our daily routine is predictable								
Age 14	1.00 (0.41)				2.25 (0.42)			

Table 5. Positive Family Relationship Scale Category Boundary Discrimination (CBD) and Intersection (Int) Parameters as Estimated Under the Nominal Response Model for Ages 9 Through 17 Years: Items 21 to 24 (R Indicates Reverse Coded Item).

Item	CBD parameters				Intersections			
	CBD ₁ (SE)	CBD ₂ (SE)	CBD ₃ (SE)	CBD ₄ (SE)	Int ₁ (SE)	Int ₂ (SE)	Int ₃ (SE)	Int ₄ (SE)
Item 21: People in our family willingly help each other								
Age 9	1.09 (0.61)	1.39 (0.83)	0.65 (0.49)		-1.67 (0.25)	-0.37 (0.45)	2.32 (0.21)	
Age 10	1.68 (0.76)	1.64 (0.56)	3.33 (1.31)		-1.37 (0.81)	-0.14 (1.07)	1.38 (0.41)	
Age 11	1.29 (0.50)	3.29 (1.31)	2.47 (1.47)		-0.89 (0.72)	0.06 (1.20)	1.34 (0.38)	
Age 12	1.28 (0.47)	0.42 (0.29)			-0.96 (0.22)	-0.62 (0.28)		
Age 13	0.60 (0.32)	0.87 (1.02)			-1.07 (0.19)	0.44 (0.22)		
Age 14	0.58 (0.40)	1.64 (2.28)	0.83 (1.41)		-1.57 (0.23)	0.24 (0.43)	1.92 (0.22)	
Age 15	0.64 (0.40)	1.65 (1.99)	0.66 (1.12)		-0.03 (0.22)	0.24 (0.41)	2.23 (0.24)	
Age 16	1.77 (0.50)				0.25 (0.34)			
Age 17	1.27 (0.50)	0.85 (0.67)	1.81 (0.84)		-0.45 (0.31)	0.25 (0.56)	1.33 (0.24)	
Item 22: It is difficult to get cooperation from family members when household chores are to be done (R)								
Age 9	0.67 (0.46)	0.20 (0.63)	0.85 (0.49)		-1.33 (0.16)	-0.70 (0.30)	0.02 (0.19)	
Age 10	0.48 (0.39)	1.10 (2.38)			-0.92 (0.22)	0.07 (0.23)		
Age 11	1.46 (0.49)	1.15 (0.37)			-1.87 (0.35)	0.98 (0.36)		
Age 12	1.78 (0.60)	0.87 (0.27)			-1.20 (0.31)	1.09 (0.33)		
Age 13	0.37 (0.15)	1.16 (1.39)	-2.62 (2.39)		-0.19 (0.17)	0.47 (0.41)	-0.33 (0.24)	
Age 14	1.06 (0.42)	0.59 (0.38)			-1.76 (0.23)	1.07 (0.26)		
Age 15	0.75 (0.38)	1.02 (1.00)	1.89 (1.86)		-0.41 (0.40)	0.44 (0.69)	1.52 (0.28)	
Age 16	1.26 (0.50)	1.15 (0.48)			-1.75 (0.29)	1.09 (0.32)		
Item 24: When there is free time, family members engage in separate activities (R)								
Age 9	1.05 (0.47)	0.22 (0.31)	0.74 (0.45)		-1.35 (0.18)	-0.50 (0.33)	1.31 (0.17)	
Age 11	0.75 (0.34)	0.29 (0.49)	0.58 (0.70)		-1.16 (0.16)	-0.10 (0.30)	2.03 (0.17)	
Age 13	0.64 (0.37)	0.82 (1.00)			-1.23 (0.19)	-0.26 (0.24)		
Age 14	1.27 (0.35)				0.34 (0.29)			
Age 15	1.45 (0.39)				0.74 (0.32)			
Age 16	1.01 (0.38)	0.91 (0.58)			-0.39 (0.22)	0.75 (0.25)

Figure 3. Category response curves for Item 3 at age 11 years modeled under the nominal response model with category boundary discrimination parameters of $a_1^* = 0.38$, $a_2^* = 0.77$, $a_3^* = 0.09$.

15 years display steeper CRCs than the lower CBD values produced at age 11 years. For age 15 years, respondents at or below 1.13 standard deviations below the mean are most likely to endorse Category 1, whereas respondents between 1.13 and 0.55 standard deviations below the mean will mostly response in Category 2. Furthermore, Category 3 is the most likely response for all respondents between 0.55 standard deviations below the mean and 2.37 standard deviations above the mean. Finally, respondents above 2.37 standard deviations above the mean are most likely to endorse Category 4. However, in Figure 4, the category distinction between response options one and two for Item 3 at age 11 years is less pronounced than at age 15 years indicating that respondents were unable to clearly distinguish between these categories. Additionally, the distinction between response Categories 3 and 4 is nonfunctioning, with Category 4 never being the most likely response category.

As denoted, the CRCs can be transformed to create the IIFs indicating the amount of information each category is providing and the overall amount of information for

Figure 4. Category response curves for Item 3 at age 15 years modeled under the nominal response model with category boundary discrimination parameters of $a_1^* = 1.00$, $a_2^* = 0.62$, $a_3^* = 0.41$.

the item. The corresponding category and item information for Item 3 for ages 11 and 15 years are displayed in Figures 5 and 6, respectively. The IIF plots contain a solid line (black in the online version), which represents the information for the overall item, the sum of the category information curves; the remaining lines indicate the amount of information provided by each category using the aforementioned line-type descriptions. For age 15 years, Category 1 provides the most information about respondents who are 1.13 standard deviations below the mean, Category 2 is the second highest peak providing information about respondents between 1.13 and 0.55 standard deviations below the mean. Additionally, Category 3 is providing information about respondents between 0.55 standard deviations below the mean and 2.37 standard deviations above the mean and respondents 2.37 standard deviations above the mean are providing information at the high end of the latent trait. The IIF curves for each category at age 11 years are indistinguishable rendering it difficult to determine the information being provided at the category level. Directly comparing the

Figure 5. Category and item information curves for Item 3 at age 11 years modeled under the nominal response model with category boundary discrimination parameters of $a_1^* = 0.38$, $a_2^* = 0.77$, $a_3^* = 0.09$.

item information for this item, it is clear that age 15 years is much more informative than at age 11 years. The IIF has a pronounced peak at 0.2 standard deviations below the mean, conversely for age 11 years, the item information is much flatter and, therefore, uninformative. Considering the CBD values, CRC, and IIF plots, this item is more appropriate and informative for 15- than for 11-year-olds and hence deemed appropriate for inclusion at age 15. In summary, this procedure was used to determine appropriate item inclusion at specific ages.

Category Functioning. To complete the scale construction process, category functioning of each item was evaluated on the revised scales to determine the appropriate response format.¹ The resulting response format was free to vary between items within each scale as well as across the 9-year span. For future use, items would be relabeled to include two (1 = *Never*, 2 = *Always*), three (1 = *Never*, 2 = *Sometimes*, 3 = *Always*), four (1 = *Never*, 2 = *Infrequently*, 3 = *Frequently*, 4 = *Always*), or five (1

Figure 6. Category and item information for Item 3 at age 15 years modeled under the nominal response model with category boundary discrimination parameters of $a_1^* = 1.00$, $a_2^* = 0.62$, $a_3^* = 0.41$.

= *Never*, 2 = *Infrequently*, 3 = *Sometimes*, 4 = *Frequently*, 5 = *Always*) response options. To illustrate, Item 22 required response format revision. As can be seen in Figure 7, the distinction between the first and second categories is functioning with a CBD of 1.46. Category 3 is redundant with Category 2, the distinction between Categories 2 and 3 is nonfunctioning with a CBD of 0.11 and does not provide unique information about PFR. The remaining response option distinction between Categories 3 and 4 function adequately with a CBD parameter of 1.13. Therefore, the nonfunctioning distinction between Categories 2 and 3 indicate that the response options for this item could be reformatted to a three-category item while maintaining psychometric information. The CBD values for the revised response option format of Item 22 (Figure 8) with three categories are 1.46 and 1.15.

Ultimately, in addition to items found informative at specific ages, 10 items were consistently informative, and thus, were included in the scale across all nine assessments, demonstrating consistency of the items across time and indicating invariance

Figure 7. Category response curves for Item 22 at age 11 years modeled under the nominal response model with category boundary discrimination parameters of $a_1^* = 1.46$, $a_2^* = 0.11$, $a_3^* = 1.13$.

of the measurement of PFR. As was evident, in the course of this longitudinal investigation, several items contributed more or less information based on the particular age. For instance, Item 3, “Input from all family members is considered when planning family activities” and Item 6, “Our family does things together rather than separately,” became informative beginning at age 13 years. Figure 9 displays the TIF for each of the nine resulting scales derived for ages 9 through 17 years. As can be seen, the information of each scale generally peaks around the mean of PFR, with the scale being informative from approximately 2 standard deviations below the mean on PFR to approximately 2 standard deviations above the mean on PFR. Additionally, the range of the information functions peak between a standardized information of approximately 0.3 for Year 14 (gray curve as presented in the online version) and approximately 0.6 for Year 16 (sky blue curve as presented in the online version). These information functions indicate that the developed PFR scale provides precise measurement of individuals across the midrange to the extremes of the PFR construct

Figure 8. Category response curves for Item 22 at age 11 years modeled under the nominal response model with category boundary discrimination parameters of $a_1^* = 1.46$, $a_2^* = 1.15$.

consistently for ages 9 through 17 years. Furthermore, the overlapping test information curves provide evidence that the scale at each year is measuring the same construct and, thus, is sufficiently invariant for the application of test equating (Brennan, 2008).

θ Scores. NRM parameter estimates for the resulting scales were used to produce nine expected a priori θ scores, which is the respondent's placement on the latent trait as interpreted on a common standardized metric. Because of the varying amount of items and response format for each assessment, the use of test equating creates a common scale of measurement for the PFR scale for the purpose of assessing validity. Having respondents on a common metric allows examination of longitudinal changes in PFR. Additionally, the use of test equating with equivalent groups can reduce the estimation error by adjusting for small differences in the measure of the latent trait between the ages (Hanson & Béguin, 2002). The process of test equating uses one or more common items as anchor items, consistently functioning across all

Figure 9. Test information of the resulting scales for ages 9 through 17 years.

test forms (Kelderman, 1988), for test calibration. The test-equating process was completed in R using the “plink” package (Weeks, 2011). This is the only package available to perform test equating using parameters estimated under the NRM. The anchor item, Item 12, was chosen based on its consistency of CBD parameter values across time and the number of response options remained constant (see Tables 1-5).

Comparison of θ Scores to Summed Scores. Table 6 presents correlations used to demonstrate the similarities and differences between the traditionally computed summed scores. The summed scores were computed by simply summing the responses across all items. Specifically, the second column contains the correlations between the original 25-item scale and the resulting θ scores from the final version of the PFR scale. As can be seen, correlations ranged between .624 and .929, more measurement error is present in the computed summed scored through the inclusion of uninformative items and nonfunctioning within-item categories. Therefore, the θ scores provide a more precise measure of the PRF construct. Furthermore, the ordering of individuals changes on the PFR construct based on the use of summed scores

Table 6. Correlations of the Final Version of the Positive Family Relationship Scale With Summed Scores Computed From the Original 25-Item Scale and the Reduced Item Scale After Uninformative Items Were Eliminated.

Year (θ scores)	Original 25 items summed score	Reduced-item summed score
9	.825*	.944*
10	.834*	.950*
11	.820*	.912*
12	.777*	.923*
13	.861*	.941*
14	.915*	.975*
15	.624*	.718*
16	.929*	.952*
17	.892*	.967*

* $p < .01$.

versus θ scores, which could be problematic if the scale is used to make decisions, particularly those that are high stakes. Importantly, the θ scores, compared with the summed score, would be expected to be more accurate in making predictions about other behavioral criteria. Therefore, it is important to use IRT to identify and remove uninformative items from a scale in order to obtain a more accurate measure of the PFR construct. However, these less than perfect correlations could have resulted by using any of the polytomous IRT models. Therefore, to evaluate the importance of using the NRM during scale development and revision, a new set of summed scores was computed from only the items found to be informative at each age. For instance, the final PFR scale at Year 9 comprised Items 1, 2, 5, 8, 9, 11, 12, 14, 17, 19, 21, 22, and 24, so a reduced-item summed score was computed by adding the responses of those items only. Specifically, the omitted Items 3, 4, 6, 7, 10, 13, 15, 16, 18, 20, 23, and 25 were not used to compute the summed score. This process was repeated for each year. The third column of the table contains the correlations between the reduced-item summed scores and the final version of the PFR θ scores. As can be seen, correlations ranged between .718 and .975, indicating that, albeit closer to a perfect correlation of one after the uninformative items were removed, there is still some difference in the ordering of individuals that can only be attributed to the modification of the response format as dictated by the NRM. Therefore, additional measurement precision is gained from using the NRM specifically.

Discussion

This article advances the field of psychological measurement in multiple ways by using the NRM under the IRT framework to construct the PFR scale. First, this is the initial application of the NRM to the construction of a scale measuring positive family relationships. Second, the scale measures a unitary construct that is customized

for each age. Third, by using the NRM, this scale contains only uniquely informative items consisting exclusively of discriminating response categories so as to optimize the accuracy of measuring positive family relationships. Fourth, this scale was constructed within a long-term longitudinal framework, and it is the first usage of the NRM with longitudinal data. Fifth, by employing longitudinal data, the study is innovative in implementing the method of test equating to a psychological construct. Last, this application of IRT and parameter linking resulted in a scale appropriate for use across the extensive developmental timeframe of middle childhood through late adolescence.

Historically, psychological scales are constructed with a fixed set of items and response format. This conventional approach restricts researchers and users of the scale by inflating measurement error due to the inclusion of unnecessary items with an arbitrary response format (Embretson & Reise, 2000). As was demonstrated during initial item analyses, several items at any given age were uninformative or redundant, requiring removal from the scale or contained nonfunctioning response categories, requiring reformatting of the response options. The removal of uninformative or redundant items produced a scale that at each age is more efficient because only necessary items are included. This would also reduce the time individuals spend responding to the scale. Additionally, through condensing the response format, cognitive load on individuals completing the scale is reduced. Individuals are no longer forced to make distinctions between arbitrary response options, rather are requested only to distinguish between meaningful response categories.

Because of the benefits of the NRM, the scale at each age contains only uniquely informative items and response category options, providing more precise measurement of the single underlying construct across all ages. As demonstrated through the overlapping TIFs displayed in Figure 9, the scale, at different ages, produced nearly identical information across similar ranges of PFR indicating the construct is unitary and is consistently measured with comparable accuracy. Therefore, the measurement of PFR has been optimized from ages 9 through 17 years. Additionally, correlations between the PFR θ scores, and the summed score computed based on the original set of 25 items, showed a reordering of individuals based on scoring. Specifically, the relative standing of individuals on the latent trait of PFR may change when using the summed score computed from original 25 items compared with the theta scores produced during the NRM estimation. Although reduced, the reordering of individuals' relative standing on the latent trait was still present even after the uninformative items were removed from the scale and a summed score was computed from the reduced items. These findings demonstrate the importance of using the unique capabilities of the NRM, identify overall item functioning in addition to within-item category functioning. As demonstrated throughout the present study, it is imperative to the overall measurement accuracy when constructing a scale that the response be meaningfully created, rather than imposing an arbitrarily defined Likert-type response format. The present study illustrates that the format of the response options should be considered equally important as is the composition of the scale items themselves. The NRM

uniquely provides the ability to conduct in-depth psychometric evaluations of scales at the category level, a necessary part of scale construction.

Because the PFR scale has the psychometric property of invariance across ages, then any changes across time can be attributed to the construct itself and not to variance in measurement. Hence, knowledge about PFR across time can be determined. Using the linked theta scores, future research can evaluate the stability with regard to interindividual differences or relative rank-order across the time span investigated, which has implications regarding predictability of PFR over time.

Although the NRM (Bock, 1972, 1997) was originally created for the application to unordered, multiple-choice tests, the present research demonstrates the utility of the NRM to ordered response formats and its ability to identify poorly functioning response options that should be eliminated. The present data set might not be considered large for the application of IRT, specifically the NRM (De Ayala & Sava-Bolesta, 1999; DeMars, 2003). However, prior studies evaluating parameter estimation under the NRM with small sample sizes used the traditional maximum likelihood estimation implemented in now outdated software. Therefore, future research is needed to evaluate the appropriate sample size for accurate estimation of NRM parameters using the supplemented EM as implemented in modern IRT software. As evidenced by the comparable size of the parameter estimate standard errors, results of the present study clearly indicate the success of using this method with such a sample size.

As with the construction of all scales, validation should be conducted. First, because the PFR was constructed within a longitudinal framework across an extensive duration, future research should determine the annual and long-term stability of the construct across childhood and adolescence. Thus, resulting PFR theta scores should prove useful in evaluating stability (interindividual differences of PFR scores across time) and continuity (mean changes in PFR scores across time) as well as their developmental meaning and significance. Second, to establish the nomological network associated with PFR, future research needs to examine its relationship to other theoretically relevant constructs. For example, PFR theta scores would be expected to relate to measures of family cohesiveness, support, interactions, and conflict, in the expected directions. PFR should also be related to a variety of children's outcomes, such as self-concept, academic performance, and educational success. Third, the relation of PFR to family socioeconomic status should be examined to determine whether these constructs are independent of each other, that is, whether or not families varying in socioeconomic status are comparable with regard to PFR. Last, the empirically derived response format should be evaluated on an independent sample of participants to further evaluate the functioning of the revised response format. In conclusion, the present psychometric research furnishes a unique approach to scale development using the NRM within a long-term longitudinal study to construct a PFR scale. The NRM provides researchers the opportunity to optimize measurement precision in developing a scale, measuring the same construct across time, to contain only uniquely informative items at each measured time point consisting exclusively

of discriminating response categories so as to optimize the accuracy of measuring positive family relationships.

Acknowledgments

We thank the parents and children for their participation in the Fullerton Longitudinal Study.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Portions of this research were supported by grants from the Spencer Foundation, Thrasher Research Fund, and California State Universities, Fullerton and Northridge.

Note

1. Whereas the Wald test was used in Preston, Reise, Cai, and Hays (2011), it was not used for revision of the response scale because prior simulation studies determined the Wald test has inadequate power to detect significant differences in category functioning for sample sizes less than 500 (Parral & Preston, 2013).

References

- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- Benson, P. L., Scales, P. C., Hamilton, S. F., & Sesma, A. (2006). Positive youth development: Theory, research, and applications. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology* (pp. 894-941). Hoboken, NJ: Wiley.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29-51.
- Bock, R. D. (1997). The nominal categories model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 33-49). New York, NY: Springer.
- Brennan, R. L. (2008). A discussion of population invariance. *Applied Psychological Measurement*, 32, 102-114.
- Cai, L. (2008). SEM of another flavor: Two new applications of the supplemented EM algorithm. *British Journal of Mathematical and Statistical Psychology*, 61, 309-329.
- Cai, L. (2012). flexMIRT: Flexible multilevel item factor analysis and test scoring [Computer software]. Seattle, WA: Vector Psychometric Group.
- De Ayala, R. J., & Sava-Bolesta, M. (1999). Item parameter recovery for the nominal response model. *Applied Psychological Measurement*, 23, 3-19.
- DeMars, C. E. (2003). Sample size and the recovery of the nominal response model item parameters. *Applied Psychological Measurement*, 27, 275-288.

- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*, Mahwah, NJ: Lawrence Erlbaum.
- Gottfried, A. E., Marcoulides, G. A., Gottfried, A. W., & Oliver, P. H. (2013). Longitudinal pathways from math intrinsic motivation and achievement to math course accomplishments and educational attainment. *Journal of Research on Educational Effectiveness*, 6, 68–92.
- Gottfried, A. W., & Gottfried, A. E. (1984). Home environment and cognitive development in young children of middle-socioeconomic-status families. In A. W. Gottfried (Ed.), *Home environment and early cognitive development: Longitudinal research* (pp. 57-115). New York, NY: Academic Press.
- Gottfried, A. W., Gottfried, A. E., Bathurst, K., Guerin, D. W., & Parramore, M. M. (2003). Socioeconomic status in children's development and family environment: Infancy through adolescence. *Socioeconomic Status, Parenting, and Child Development*, 287, 189-207.
- Guerin, D. W., Gottfried, A. G., Oliver, P. H., & Thomas, C. W. (2003) *Temperament: Infancy through adolescence: The Fullerton Longitudinal Study*. New York, NY: Kluwer Academic.
- Hanson, B. A., & Béguin, A. A. (2002). Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the common-item equating design. *Applied Psychological Measurement*, 26, 3-24.
- Harvey, R. J., & Hammer, A. L. (1999). Item response theory. *The Counseling Psychologist*, 27, 353-383.
- Hollingshead, A. B. (1975). *Four factor index of social status*. Unpublished manuscript, Yale University, Department of Sociology, New Haven, CT.
- Kelderman, H. (1988). Common item equating using the loglinear Rasch model. *Journal of Educational Statistics*, 13, 319-336. doi:10.2307/1164707
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16, 159-176.
- Ostini, R., & Nering, M. L. (2006). *Polytomous item response theory models*. Thousand Oaks, CA: Sage.
- Paradis, A. D., Giaconia, R. M., Reinherz, H. Z., Beardslee, W. R., Ward, K. E., & Fitzmaurice, G. M. (2011). Adolescent family factors promoting healthy adult functioning: A longitudinal community study. *Child and Adolescent Mental Health*, 16, 30-37. doi: 10.1111/j.1475-3588.2010.00577
- Parral, S. N., & Preston, K. S. J. (2013). *Evaluating the Wald test under nonnormal conditions using empirical histograms*. Poster session presented at the meeting of the American Psychological Association, Honolulu, HI.
- Preston, K. S. J., & Reise, S. P. (2014). Detecting faulty within-item category functioning with the Nominal Response Model. In S. P. Reise & D. Revicki (Eds.), *Handbook of item response theory modelling: Applications to typical performance assessment* (pp. 386-405). London, England: Routledge.
- Preston, K. S. J., Reise, S. P., Cai, L., & Hays, R. D. (2011). Using the nominal response model to evaluate response category discrimination in the PROMIS emotional distress item pools. *Educational and Psychological Measurement*, 71, 523-550.
- R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Reise, S. P. (2009). Item response theory and clinical measurement. *Clinical Psychology*, 5, 27-48.

- Reise, S. P., Ainsworth, A. T., & Haviland, M. G. (2005). Item response theory: Fundamentals, applications, and promise in psychological research. *Current Directions in Psychological Science*, 14, 95-101. doi:10.1111/j.0963-7214.2005.00342.
- Reise, S. P., & Henson, J. M. (2003). A discussion of modern versus traditional psychometrics as applied to personality assessment scales. *Journal of Personality Assessment*, 81, 93-103. doi:10.1207/S15327752JPA8102_01.
- Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology*, 5, 27-48.
- Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores* (Psychometric Monograph No. 17). Richmond, VA: Psychometric Society. Retrieved from <https://www.psychometricsociety.org/sites/default/files/pdf/MN17.pdf>
- Sy, S. R., Gottfried, A. W., & Gottfried, A. (2013). A transactional model of parental involvement and children's achievement from early childhood through adolescence. *Parenting: Science and Practice*, 13, 133-152. doi:10.1080/15295192.2012.709155.
- Thissen, D., Steinberg, L., & Fitzpatrick, A. R. (1989). Multiple-choice models: The distractors are also part of the item. *Journal of Educational Measurement*, 26, 161-176.
- Ura, M., Preston, K. S. J., & Mearns, J. (2014). A measure of prejudice against accented English (MPAAE): Scale development and validation. *Journal of Language and Social Psychology*. Manuscript submitted for publication.
- Weeks, J. P. (2011). Plink package [Computer software]. Retrieved from <http://cran.r-project.org/web/packages/plink/index.html>