Constructing and Evaluating Measures: Applications of the Rasch Measurement Model

Kelly D. Bradley[1], Jessica Cunningham, R. Trent Haines, W. E. Harris, Jr., Christian E. Mueller, Kenneth D. Royal, Shannon O. Sampson, Gilbert Singletary and Jennifer A. Weber[2]

[1] Use Kelly D. Bradley at author of contact. University of Kentucky, Department of Educational Policy and Evaluation Studies. 131 Taylor Education Building, Lexington, KY 40506-0001.
email: kdbrad2@uky.edu phone: (859)257-4923

[2] All authors contributed equally to this paper. Authors are listed in alphabetical order.

Constructing and Evaluating Measures: Applications of the Rasch Measurement Model

Abstract

This paper is a collaboration of student-lead research projects applying the Rasch measurement model with Winsteps software. Various applications, including higher education, educational psychology and counseling, demonstrate the theoretical framework, methodological process and results utilizing Rating Scale and Assessment formats. The basics of Rasch measurement are overviewed [including a comparison to Classical Test Theory], followed by specific applications, including background, specific methods, and a discussion of the results and educational contributions. In addition, sample code is presented so that the examples can be modeled by the readers. The paper demonstrates the strengths of applying Rasch measurement across various fields and will provide a foundation for others to utilize the model as a methodological tool.

Constructing and Evaluating Measures: Applications of the Rasch Measurement Model

Rasch measurement is potentially relevant whenever an assessment or questionnaire is constructed to measure the degree of some property inherent in persons or other entities. The Rasch model is a one-parameter logistic model within item response theory (IRT) in which the amount of a given latent trait in a person and the amount of that same latent trait reflected in various items can be estimated independently yet still compared explicitly to one another. The purpose of this paper is to introduce participants and readers to the theory and a collection of applications of Rasch measurement and provide real-world examples via Winsteps software.

An introduction to the Rasch model

The Rasch model is the simplest of the IRT models, having a single parameter for the person or entity and a single parameter corresponding to each category of an item. A Rasch analysis is appropriate any time a researcher wishes to use the total score on an assessment or questionnaire to make inferences about an individual's ability or level of a latent trait inherent in that individual. Support for this conjecture comes from the fact that the Rasch model follows mathematically from the requirement of invariance of comparisons among persons and items. While it is the case that Classical Test Theory (CTT) also uses the total score to characterize each person, CTT asserts the total score as the relevant statistic with little consideration of anomalies in the items or the individuals answering them. Applications of the Rasch model account for these anomalies and provide a more informative score. The aim of Rasch measurement is analogous to the construction of a ruler, establishing the correct measure (Andrich & Luo, 2003).

The Rasch model, introduced by Georg Rasch (1980), yields a comprehensive and informative picture of the construct under measurement as well as the respondents on that

measure. It allows observations of respondents and items to be connected in a way that indicates the occurrence of a certain response as probability rather than certainty and maintains order in that the probability of providing a certain response defines an order of respondents and items. In other words, a person endorsing an extreme statement, or answering a difficult item, should also endorse all less extreme statements, or answer correctly the less difficult items (Wright and Masters, 1982). It is a static model, in contrast to a sequential processing model, meaning that for each person having a certain amount of the given latent trait; it specifies the probability of a response in one of the categories of an item. The response process is a classification into ordered categories defined by thresholds, the point where the probability of a response in either one of two adjacent categories is 50%. This again is analogous to markings on a ruler, with the exception that these markings are estimates (Andrich & Luo, 2003; Guttman, 1950).

The Rasch model uses the sum of the item ratings simply as a starting point for estimating probabilities of those responding to each item. In the case of a questionnaire, it is based upon an individual's willingness to endorse a set of items and the difficulty to endorse those items. In the case of an assessment, it is based on an individual's ability on a construct and the difficulty of the items on the assessment. In both cases, it is assumed item difficulty, or difficulty to endorse, is the main characteristic influencing responses (Linacre, 1999). Rasch analysis reports both person ability or willingness-to-endorse and item difficulty or difficulty-to-endorse estimates along one logit (log odds unit) scale, "a unit interval scale in which the unit intervals between the locations on the [combined person-item scale] have a consistent value or meaning" (Bond and Fox, 2001, p. 29). Bond and Fox explain that employing Rasch techniques allows for the ordering of respondents along this continuum of ability or willingness to endorse

items, and orders items along a continuum of difficulty (to endorse). The examples in this paper utilize Winsteps, a Rasch modeling software program available at www.winsteps.com.

This paper is a collaboration of student-lead research projects applying the Rasch measurement model with Winsteps software. Various applications including higher education, educational psychology and counseling demonstrate the theoretical framework, methodological process and results utilizing rating scale and assessment formats. Each example provides a brief background and/or theoretical framework, an overview of the specific methodology (including sample code in the appendix), abridged results, a discussion and the educational contribution.

Rasch Model Applications

The research collaboration outlined in this paper provides an example of cross-disciplinary efforts sharing a methodological tool. Various applications of the Rasch model are presented in an effort to introduce participants and readers to the theory and applications of Rasch measurement applied to assessments and questionnaires. The contribution serves as a teaching tool, while demonstrating the strengths of Rasch measurement across various fields. Seven examples are presented:

Example 1	Tacit Knowledge of Superintendents	Page 5
Example 2	Traditional Religious Principles and Objectives	Page 9
Example 3	American Association for Employment in Education Survey	Page 13
Example 4	Black University Students' Perceptions of Marriage	Page 19
Example 5	Community College Faculty Perceptions of Student Outcomes	Page 24
Example 6	Therapist Characteristics	Page 28
Example 7	Psychological Sense of School Membership Scale	Page 33

Example 1 – ***Exploring Construct Validity of Tacit Knowledge in a Sample of Kentucky Superintendents: Employing the Partial-Credit Rasch Model***

Christian E. Mueller; University of Memphis

Tacit knowledge is best characterized as a form of procedural knowledge gained through experience, rather than through formal training or instruction. Although Wagner (1987) found empirical support for a single underlying construct of tacit knowledge, most researchers (e.g. Colonia-Wilner, 1999; Edwards & Schleicher, 2004) view tacit knowledge in three areas: intrapersonal, interpersonal and organizational. Intrapersonal tacit knowledge refers to being able to prioritize work tasks efficiently and effectively. Interpersonal tacit knowledge refers to knowing how best to manage or lead others. Organizational tacit knowledge is concerned with being able to successfully accomplish the larger goals of the organization (see Sternberg et al., 2000). The interest in tacit knowledge as a research construct continues to grow (e.g. Nestor-Baker & Hoy, 2001; Tschannen-Moran & Nestor-Baker, 2004).

Objective

The example presented was a component of a larger effort to develop the Tacit Knowledge Inventory for School Superintendents (TKIS), which in its final form will be a scenario-based assessment that measures intrapersonal, interpersonal, organizational tacit knowledge. Here, the goal was to provide initial construct validity for items developed for use on the three dimensions.

Method

Sample

Participants were selected from the final piloting phase of the larger study. The sample consisted of two groups: graduate students enrolled in a superintendent preparation program in

the state of Kentucky and superintendents currently practicing in the state of Kentucky. The groups represented two levels of experience, novice and expert, which is necessary in examining differences in levels of tacit knowledge. In total, six graduate students and 29 superintendents returned a completed inventory, resulting in 35 responses for the Rasch analysis.

Instrumentation

A draft form of the TKIS was distributed to practicing respondents via mail and to graduate students by classroom instructor. The draft form of the TKIS consisted of eight scenarios that represented typical day-to-day situations faced by superintendents. In addition, each scenario was accompanied by up to five possible ways of handling the main problem presented. Respondents were asked to rate each of the response options utilizing a 7-point scale where 1 = extremely ineffective to 7 = extremely effective. Point totals were assigned according to how well each response matched the scoring protocol, and ranged from 0 (worst) to 3 (best). The scoring protocol was developed from the mean responses of a pre-identified expert panel, defined here as superintendents with at least 10 years of experience who were considered by reputation as being particularly effective leaders (Fox, 1999; Sternberg et al., 2000).

Data Analysis

The partial-credit Rasch model (Masters, 1982) is an extension of the dichotomous model and is used for polytomously scored items, such as the case with items on the TKIS. The partial-credit Rasch model is expressed as $ln(P_{nij}/P_{ni(j-1)}) = B_n - D_i - F_{ij} = B_n - D_{ij}$, where F_j is the "calibration" measure of category j relative to category $j-1$, the point where categories j and $j-1$ are equally probable relative to the measure of the item (Linacre, 2005, p. 20). Fox (1999) suggests that the partial-credit model is particularly useful in developing assessments of complex cognitive ability where "graded judgments of performance can be scaled and subsequently

assessed for quality of calibrations" (p. 342). This is indeed the case with the TKIS.

Response data were analyzed separately for the intrapersonal, interpersonal, and organizational tacit knowledge measures using Winsteps software (Linacre, 2004, version 3.51). Infit and outfit statistics and variable maps provided the basis for determining how well the items measured each type of tacit knowledge. The theoretical framework supported this inquiry.

Results and Discussion

A total of four items representing one scenario were analyzed in relation to intrapersonal tacit knowledge, 17 items representing four scenarios were analyzed in relation to interpersonal tacit knowledge, and 13 items representing three scenarios were analyzed in relation to organizational tacit knowledge. Given the methodological focus, results related to organizational tacit knowledge are presented in this paper, as it was deemed the most illustrative example.

Organizational Tacit Knowledge

Table 1 illustrates that all items included on the organizational tacit knowledge measure fit the expectations of the Rasch model. Specifically, the items fall within the acceptable infit and outfit limits of .6 to 1.5. Fox (1999) utilized these same cutoff values in her development of the Practical Knowledge Inventory for Nurses, which is a similar assessment.

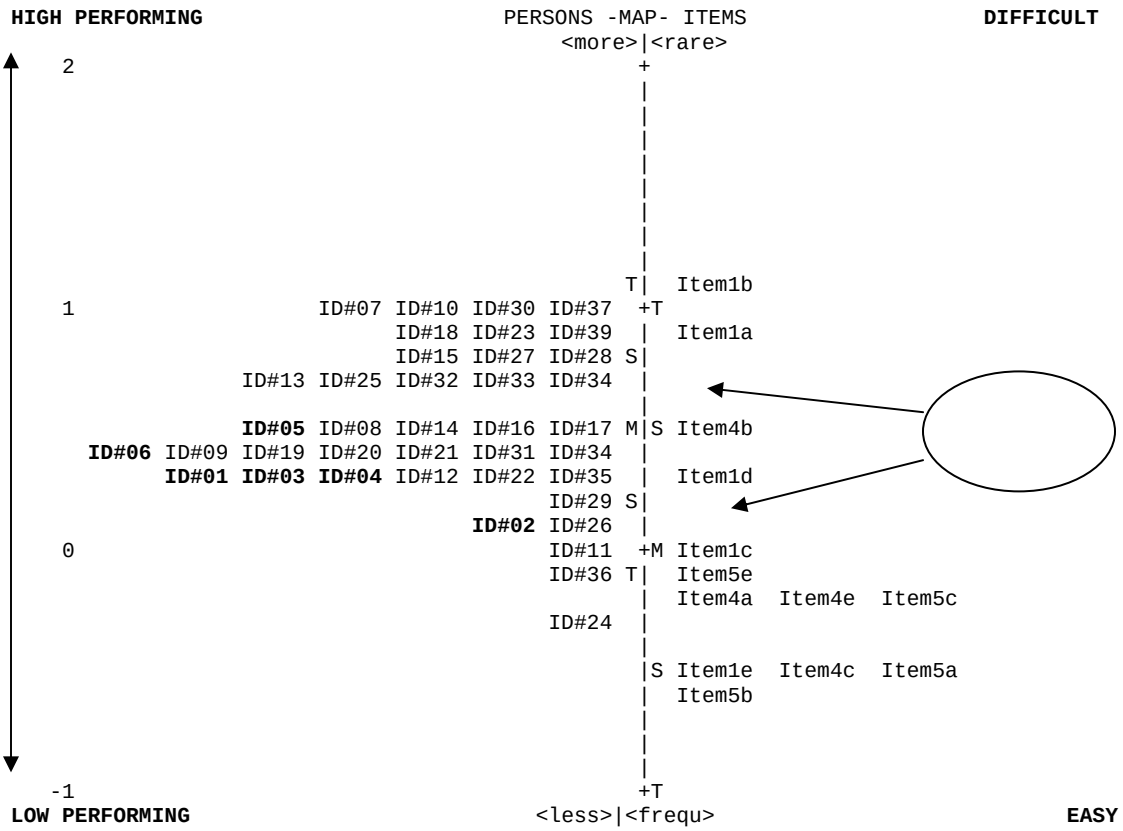
Table 1. *Fit Statistics for Organizational Tacit Knowledge Phase III (abbreviated)*

Item	Infit MNSQ	Outfit MNSQ
5a	1.41	1.41
1c	1.26	1.26
1b	1.16	1.16
4e	.87	.79
4c	.77	.61
Mean	1.00	1.01
S. D.	.12	.19

In examining the variable map, Figure 1, for organizational tacit knowledge, several things become evident. The items on the organizational tacit knowledge measure appear to have sufficient distribution, measuring at various levels of the variable. Even so, there are a few gaps in the distribution, as indicated by the ellipse with arrows in Figure 1. Given the variability of items, as evidenced by a wide distribution over the continuum (spanning 2 standard deviations, (S and T) above and below the mean (M)), there is support for sufficient item targeting. Thus, the most able respondents on this measure are being adequately measured by this bank of items.

Lastly, in comparing the performance of the graduate students to that of the practicing superintendents, Figure 1 indicates that practicing superintendents, in general, are performing at a higher level than the graduate students. Whether this is a reflection of the sample or the nature of tacit knowledge would require further empirical study to determine.

Figure 1. Variable (person/item) map for organizational tacit knowledge



The results presented here demonstrate that items generated to measure organizational tacit knowledge fit the expectations of the Rasch model and reflected the theoretical framework demonstrated in various literature. While these results do not alone prove that tacit knowledge exists in superintendents as purported by other researchers (e.g. Nestor-Baker & Hoy, 2001), the results of the present study do offer support for continuing research in this area.

Educational Significance

If tacit knowledge does indeed guide expert behavior, then there is a need for assessments that measure tacit knowledge in a range of fields in order to provide researchers with the tools necessary to understand the construct. The results presented here provide support regarding the construct validity of tacit knowledge on this assessment for researchers of superintendent tacit knowledge. By continuing such endeavors, researchers may one day be able to help individuals learn the skills necessary to acquire and use tacit knowledge in their day-to-day decision-making.

Example 2 – *An Exploratory Examination of Traditional Religious Principles and Objectivity*

Gilbert Singletary; University of Kentucky

Perhaps one of the most proverbial sentiments echoed by traditional clergy, ministers, and members of traditional Christian churches is “no sin is greater than the other.” Such language of reference is found in the Christian Bible (Rom 6:23, Rom 2:12, John 3:4). Despite the Biblical claim, the question arises as to whether or not clergy and their congregations indeed view various sins as equally severe.

Objective

The purpose of the full study was to assess clergy and affiliated church members’ perceptions regarding traditional religious principles as are prescribed in the Biblical text. In this

example, an examination of the item hierarchy related to the theoretical expectation of equivalence of sins is reviewed to verify or dispute adherence to the text.

Method

A target sample of 30 African American Baptist church members participated in the study. Participants included both males and females, ranging in age from 18 to 60. Data were collected using the Religious Attitude Survey (RAS) which consists of 32 single word items on which participants were asked to rate their feeling utilizing a 11-point scale with a range of negative five to positive five, with zero serving as a neutral point. Utilizing data displayed in the religiosity and spirituality literature, the RAS scale was developed with the Ten Commandments as its foundation to ascertain distinctions in perceptions of severity of sin. The Rasch Rating Scale Model, via Winsteps software, was implemented due to its capacity to create a measure for a single attribute from a collection of items related to that attribute. An item/person map was constructed to illustrate the empirical hierarchy of items on the survey, as connected to the participant's level of ability to endorse each item. The items listed at the lower end of the map represent a higher probability of being endorsed, as opposed to those at the top representing a lower probability of being endorsed. Participants follow a similar pattern. For example, those at the lower end of the map represent less willingness to favorably endorse an item.

Results and Discussion

Given the suggested theory "no sin is greater than the other", one would expect the items connected to sins to cluster together with little variation (if any) in difficulty to endorse. On the other hand, items that are connected to positive acts of religion should cluster together, again with little variation in endorsability level. Thus, a dichotomous distribution should be present.

This is not the case, as the empirical hierarchy does not match the theoretical expectation, at least in the context of sins.

The empirical hierarchy (presented below in Figure 1) does appear to separate out the positive acts of religion from the sins, as evident from items like Love, Forgiveness Praying and Fellowship clustering together. Even so, the same cluster is not apparent for the rating of sins. Figure 1 clearly displays some items are easier to endorse as *bad*, including Child Abuse, Murder and Rape. These items are almost two standard deviations below such items as Gambling, Cursing and Pre-marital sex, all of which are recognized sins within this congregation. In general, it appears that a value system is embedded within the participants' rating structure and there is not an agreement that *all sins are equal*.

Figure 1. *Item Map*

```
INPUT: 30 PARTS, 33 ATTS  MEASURED: 30 PARTS, 33 ATTS, 11 CATS     WINSTEPS 3.51
--------------------------------------------------------------------------------

         <more>||<frequ>
    3          ++
               ||
               ||T
               ||
               ||  23Love
               ||
    2          ++  42Forgiveness
               ||  10Salvation     18Praying     22Grace
               ||  13Witness     Fellowship
               ||
               ||S
               ||
    1          ++
               ||
               ||
               ||
            X  ||
            X T||  26Gay Rights     31Gambling     33Cursing
    0      XX  ++M 20Pre-Marital sex   27Drunkenness  34 Same sex Adp     41Gay parents
          XXX S||  29Same Sex Coup     30Lusting     36Lying     39Fornication
        XXXXX  ||  12Bi-Sexual     25Lesbian
      XXXXXXX M||  19Gay     21Homophobia     32Transexual     40Greed
         XXXX  ||  15Homosexuality     38Prostitution
         XXXX S||  16Adultery     24Stealing     28Worshiping other Gods
   -1      XX  ++
            X T||
               ||S
               ||
               ||
               ||
   -2          ++
               ||
               ||
               ||
               ||T
               ||
   -3          ++  37Rape
               ||
               ||
               ||
               ||  14Murder
               ||
   -4          ++
               ||
               ||
               ||
               ||
               ||
   -5          ++  11Child Abuse     35Child Sex abuse
         <less>||<rare>
```

Educational Significance

The majority of participants did not maintain the position that all sins are judged equal, as is often assumed and asserted by leaders of the church. Instead, it appears that a hierarchy does

indeed exist. This study provides a foundation of empirical evidence to support the contention that the religious community does indeed view sins in a hierarchical format, not the clear picture of right and wrong as is often presented. This finding has implications for church practices and the way educational researchers set out to measure such dispositions.

Example 3 – *Examining Specifications for the Partial Credit Model with the AAEE Survey*

Jessica Cunningham; University of Kentucky

For nearly three decades, the American Association for Employment in Education (AAEE) has conducted a survey of higher education institutions' perceptions on teacher supply and demand. Results from the AAEE yearly report have received national attention. The results have an inherent purpose in data-driven policy decisions for k-12 and teacher education settings.

Objective

This example addresses the issue of portraying an accurate hierarchical structure through examining item calibrations for the items on the AAEE survey related to perceptions of the need for educators in various fields. The hierarchical structure of these items is illustrated with the specification of the person ability center set to zero and compared to the results based on the Winsteps default specification, in which the item difficulty center is set to zero. In addition, the issue of interpreting the results under each scenario is considered. Specifically, the following research questions are posed:

- What effect does setting the mean of the person abilities to zero for the partial credit model have on the hierarchical structure of the demand for teaching fields?
- How does this analysis impact the interpretation of the results?

Determining proper specifications for the partial credit model is essential to conducting accurate analysis. The purpose of this study is to address the issue of properly applying the partial credit model to provide a meaningful hierarchical structure of the perceived demand in teaching fields from the AAEE annual survey.

Method

Responses for the AAEE Educator Supply and Demand survey are solicited from all institutions preparing educators, as listed in the Higher Education Directory (HED), which includes AAEE members and nonmembers. Surveys are sent to either the institutions' directors of career services or deans/directors of teacher education divisions.

Instrumentation

Participants may complete the survey either online or in a pencil-and-paper format, with the majority of respondents opting for the latter option (AAEE, 2004). The survey instrument, which is mailed during the Spring semester, is comprised of a number of sections relating to educator supply and demand; however, the focus here is on the teaching fields section. For 64 teaching fields, respondents are asked to utilize a Likert-type scale to rate the perceived demand for teaching fields offered by their institution. The rating scale ranges from 1 = considerable surplus to 5 = considerable shortage, with a neutral point of 3 = balance. The AAEE-constructed scale is: 4.21-5.00 = considerable shortage; 3.41-4.20 = some shortage; 2.61-3.40 = balanced; 1.81-2.60 = some surplus and 1.00-1.80 = considerable surplus (AAEE, 2004).

Procedures for Analysis

The underlying assumption of the Rasch rating scale model is the concept of measuring only one attribute at a time, or unidimensionality (Bond & Fox, 2001). In the case of the AAEE educator supply and demand survey, the component of the instrument utilized in this study asks

respondents to rate their perceived demand for educators by each teaching field. These survey items only measure the attribute of perceived demand for educators by each teaching field. Likert-type scales may assign number labels, similar to the perceived demand options 1 to 5, but the assignment of these numerals does not equate to equidistant categories (Smith & Smith, 2004). The Rasch partial credit model employed here transforms ordinal, response rating data into equal interval scales, given the data fit the model. Mathematically, the partial credit model is represented by log $(P_{nik} / P_{ni(k-1)}) / B_n - D_i - F_{ik}$, with the scale calibration represented by F_{ik} where the rating scale threshold for item *i* is the point of equal probability of adjacent categories *k-1* and *k* (Wright & Masters, 1982).

Interpretation of the AAEE survey data is concerned with the location of the various teaching fields along a shortage continuum. For this reason, a specification for the person center, or mean, to be located at zero was entered into the Winsteps (version 3.57.1) code. For comparison, an analysis was conducted and examined with and without this specification. To determine measure stability and accuracy, the review begins with fit statistics to assess whether the assumption of unidimensionality holds empirically (Andrich, 2004; Linacre, 2004; Bond & Fox, 2001). OUTFIT mean-square fit statistics (MNSQs) are equivalent to a chi-square statistic; values greater than 2.0 indicate unexplained randomness throughout the data (Linacre, 2004; Smith, 1996). Item estimates of the construct being investigated are provided along a logit (log odds) continuum with error estimates for each item. Probability estimates indicate the likelihood for items to be perceived at a particular level of demand based on where the item estimate falls along the continuum relative to the person difficulty rating estimate, allowing each teaching field to be examined so that a hierarchical structure of the teaching fields can be formed. Item maps

from each analysis are used for comparison to determine how the person center specification impacts the item calibrations.

Results and Discussion

Responses from 360 teacher education institutes across the nation were included in this analysis. Prior to interpreting item maps, fit statistics for the teaching fields were examined. Changing the center specification does not change the fit statistics; therefore, the fit statistics in either setting may be reviewed to assess model fit. All 64 teaching fields had MNSQ fit statistics less than 2; therefore, fit of the data to the model is acceptable. The item maps with zero item center, see Figure 1, and zero person center, see Figure 2, are presented below. Item maps visually layout the hierarchical structure of the 64 teaching fields based on empirical evidence. The teaching fields are arranged in a hierarchy from teaching fields not often perceived in shortage (seen at the bottom of the figures) to teaching fields often perceived in shortage (seen at the top of the figures).

Figure 1. *Teacher Education Institutes Map of Perceived Demands Item Center Zero*

```
 <perceived more>|<demand>
6             .  +
              .  |
                 |
                 |
              .  |
5                +
                 |
              .  |
                 |
                 |
4                +
              .  |
                 |
              #  |
                 |
3             .  +
                T|
             .#  |
             .#  |
            .##  |
2             .  +T SOC
            .## S|  HEA  PE
          #####  |  PRIM
         .#####  |  ART  JOU  K   PRE
      .########  |  DAN  INT  THT
1       #######  +S BUS  DRT
     ########## M|  ENG
     .#########  |  BSM CL CUR GER GIF HU JAP MI MUSG MUSV MUSI SPE SSO
      .########  |  COU  FREN
      .########  |  AG  CS  HOM  LIB  REA
0          .###  +M PRINEL
         .##### S|  OCC  PRINMID  PRINHS  SPS  SUP  TECH
            .##  |  SNU
             .#  |  AUD  BIO  ESL  GEN
              .  |  EAR  SPA
-1            . T+S PHYT  SPP
              .  |  BILING  CHEM  EAR
              .  |  DUAL  HEAR  MENT  PHYS
              .  |  LD  MAT  MMD  VIS
              .  |  EMO  MUL  SPD
-2               +T
 <perceived less>|<demand>
```

Figure 2. *Teacher Education Institutes Map of Perceived Demands Person Center Zero*

```
 <perceived more>|<demand>
5              .  +
               .  |
                  |
                  |
               .  |
4                 +
               .  |
                  |
                  |
                  |
3              .  +
                  |
               .  |
               .  |
               .  |
2                T+
              #   |
             .#   |
             ##   |
             .#   |T
1            .# S+  HEA  PE   SOC
          .#####  |  PRIM
            .###  |  ART  JOU  K
      .#########  |  INT  PRE  THT
        .#######  |S DAN
0    .######### M+  BUS  DRT  ENG
     ##########  |  CL  CUR  GER  GIF  HU  JAP  MID  MUSG  MUSV  SSO
      .########  |  BSM  FRE  MUSI  SPE
       ########  |  AG   COU  CS   LIB
           .###  |M HOM  PRINEL  REA
-1        ###### S+  OCC  PRINMID  PRINHS  SPS  SUP  TECH
             ##  |  SNU
             .#  |  AUD  BIO  GEN
              .  |  EAR  ESL
              .  |S PHYT  SPA  SPP
-2            . T+  BILING  CHEM
              .  |  EAR  HEAR  MENT  PHYS
              .  |  DUA  LD   MAT  VIS
              .  |  EMO  MMD  MUL  SPD
                 |
-3               +T
  <perceived less>|<demand>
```

Both item maps illustrate the average person measure is above the average item measure, indicating the overall tendency for respondents rate fields as being perceived in shortage. The teaching fields perceived to be in the least amount of demand are health, physical education, and social studies. Teaching fields perceived to be in the highest demand are related to special education: emotional/behavior disorders, mild/moderate disabilities, multicategorical, and severe/profound disabilities (see Figure 2).

In Figure 2, where the person center is at zero, it is easy to identify items below zero as fields with more respondents perceiving a shortage. Teaching fields following this description include many of expected fields including special education, science and math. In contrast, Figure 1 may lead to false interpretations, such as that teaching fields at zero are perceived to be in balance. In both item maps, teaching fields fall into three clusters, relatively. Notice in both item maps, the teaching fields in the lowest cluster correspond to fields often labeled in the literature as fields in considerable shortage. Therefore, results from this study confirm shortage fields mentioned in the literature.

Educational Significance

Accurately determining the hierarchical structure of teaching field demand has policy implications for teacher education programs and also local, state and federal governments. Graphical representations present a clear picture of the structure of the teaching fields, but the responsibility exists to provide an accurate picture of the data so that no misconceptions surface when the item map is examined. Researchers can use these results to accurately model and present interpretable information for other similar survey instruments.

Example 4 – *A Study of Black University Students' Perceptions of Marriage*

William E. Harris, Jr.; University of Kentucky

Over the past decade, the institution of marriage has been influenced by factors such as divorce, infidelity and financial strain, with the Black community surfacing as one of the most affected subgroups. Compared to other populations, Black marriages have the fastest declining rate (Lawson, 2000; Taylor, 1998). Broman (1993) suggests this decline is related to expectations brought to the marriage consisting of financial support, emotional support, sexual

behavior and childcare. Other factors may include the increasing rate of incarceration among black males (Kiecolt & Fossett, 1997) and the unequal sex ratio as reported by the 2000 U.S. Census. Others suggest drug problems, excessive rates of unemployment and the growing AIDS epidemic may be to blame (Bethea, 1995; Taylor, 1998; King, 1999; Kinnon, 2003).

Objective

This example provides a review of the early process of validation of the instrument used to collect black university students' perceptions of marriage. Each item is reviewed according to its fit and function within the framework of the assumptions of Rasch analysis. Results are intended to support the revision of the instrument.

Method

The sample consists of black male students, ranging in age from 18 to 30, who were members of black student associations at a southeastern University. The study operationalizes "black" to encompass all students who identify themselves racially as Black, African American, or Bi-Racial (having one black parent). In total, 25 responses were included in the analysis.

Instrumentation

The statements on the selected-response, pencil-and-paper survey were constructed using a thorough review of the literature, as well as prevalent cultural trends. The survey consists of 20 statements, such as

- *For a black man, there are more benefits of marriage than shortcomings.*
- *Married men spend less time with their friends.*

Respondents were asked to rate their agreement utilizing a 4-point Likert-type scale where 1 = strongly disagree, 2 = disagree, 3 = agree and 4 = strongly agree.

Data Analysis

Data were analyzed using Winsteps software. Code was written to represent each respondent and Likert-type survey statement. The Rasch model permits item difficulty for each statement presented in the survey derived by the way apt participants actually respond to the statements (Bond and Fox, 2001). Variable maps illustrate an item hierarchy that identifies the level of endorsability (or difficulty) of the statements. Probability curves were used to identify if various categories were used and functioned as expected.

Results and Discussion

Figure 1 presents a map of the items, ranked by level of difficulty to endorse, and the respondents, ranked by their willingness to endorse the items. Within the map, items have been labeled by a key word in the statement. Items that are located at the top of the map have been identified as those that are more difficult to endorse. Those at the bottom are easier to endorse; thus, as you move from bottom to top of the map, items are more difficult to endorse. Results suggest the easiest item to endorse is item 8, *Your level of education can make you a more "attractive" partner when it comes to marriage*. The most difficult to endorse is item 2, *Black men do not have to work on "maintaining" the marriage because they can always find another woman*. The average ability of the respondents, which is conceptualized as respondents' willingness to endorse, is in close approximation to the average difficulty of the statements, so the items are well targeted to the group. More so, there is a general spread of items, but there is a clear gap, indicated in the figure by a bold ellipse, where items are not tapping respondents' willingness to endorse items. Attention should be given to making sure the hierarchy is complete and representative of what is presented in the literature.

Figure 1. *Item/Respondent Map*

```
                        MALES -MAP- PERS
                            <more>|<rare>
 2                                +T
                                  |  2maintain
                                  |
                                  |  10fear        17want
                                  |  7faithful
                              19 T|
 1                                +S 9divorce
                         20  22  |
                          8  14 S|
                 1   9  17  23  |
                                  |  5sexinc
    5   6   7  10  12  24  25 M|  13infedility
 0                        4  11  +M 16goodjob     4goodbw
                         15  21  |  12triflin     15betterjob   6slavery
                          2  18 S|  11monogamy
                                  |  14lesstime    19sexdec
                          3  13  |
                                 T|  18lockdown    20sucinsoc
-1                                +S
                              16  |  1benefits     3head
                                  |
                                  |
                                  |
                                  |  8edattrac
-2                                +T
                            <less>|<frequ>
```

Figure 2 illustrates a category probability curve which focuses on the rating scale configuration, specifically related to use. Peaking response categories illustrate the usage of that particular response category. Based on the probability curves, those responding are utilizing each response category of the instrument, providing comparable opportunity for selection.

Figure 2. *Category Probability Curve*

```
        CATEGORY PROBABILITIES: MODES - Structure measures at intersections
P      ++------+------+------+------+------+------+------+------++
R  1.0 +                                                          +
O      |                                                          |
B      |                                                       44|
A      |11                                                   44  |
B   .8 +  11                                               444    +
I      |    11                                            4       |
L      |      1                                         44        |
I      |       11          222222                     4          |
T   .6 +         1      222      22                 44           +
Y      |          11  22           22              4             |
    .5 +            12               2   3333333  4              +
O      |           221                2*3       3*               |
F   .4 +         22   11             33 2       4 33             +
       |        2       1           3    2    44    33           |
R      |      22         11       33      2  4        33         |
E      |    22             1    33         2*           33       |
S   .2 +  22                1133          44 22           33     +
P      |22                  3311        44     2            333  |
O      |                 333    111   44        222            33|
N      |            333333        44***11            22222        |
S   .0 +***********44444444444444       111111111111111*********+
E      ++------+------+------+------+------+------+------+------++
       -4     -3     -2     -1      0      1      2      3      4
          MALE [MINUS] PER MEASURE
```

With regard to whether this instrument is a valid tool to utilize for data collection, the results provide preliminary feedback. Each statement should play a significant part in the way a construct is being investigated. In other words, do the statements/items “fit” the construct? Table 1 reflects fit-order statistics presenting items that appear to be influenced by outside factors (outfit) and those that display “off-variable noise” (infit). When addressing infit and outfit, a mean squared value range cutoff is determined by the size of the sample. Being that the sample used in this study is less than 500 participants, the acceptable mean squared outfit range has been set as 0 to 1.3 (Bond and Fox, 2001). Outfit mean square statistics for items 2 and 7 fall outside the indicated range, suggesting these statements either are not supporting the underlying construct or the items need revamped, likely because respondents are viewing the items differently than that intended by the researcher.

Table 1. *Item Statistics in Order of Misfit (abbreviated)*

Item		INFIT		OUTFIT		
	Model S.E.	MNSQ	ZSTD	MNSQ	ZSTD	Item Description
7	.34	**1.83**	**2.3**	**1.80**	**2.2**	faithful
2	.34	**1.77**	**2.4**	**1.73**	**2.4**	maintain
3	.28	**1.31**	**1.3**	1.30	1.3	head

Educational Significance

With regard to survey construction, the Rasch model is "intuitively more satisfactory and mathematically more justifiable" (p. 71) than the traditional statistical procedure granting all items the same value when using Likert-type scale data (Bond & Fox, 2001). A good Likert-type instrument is grounded in items with a varying degree of difficulty to assess a range of attitudes held by the participants. Using Rasch analysis in the preparation stages of an instrument will result in a more sound measure, leading to meaningful results.

Example 5 – *Using Differential Item Functioning (DIF) to Measure Community College Faculty Perceptions of Student Outcomes*

Kenneth D. Royal; University of Kentucky

Previous research suggests differences exist between the academic sciences. Much research suggests faculty from the *hard* sciences, such as math, physics and engineering, are concerned primarily with the intellectual development of students and less concerned with the noncognitive factors like social and moral growth (Smart, Feldman and Ethington, 2000). In contrast, it has been argued that faculty from the social and behavioral sciences are largely concerned with all domains, even emphasizing the noncognitive factors.

Objective

The purpose of this study is twofold: 1) to demonstrate differential item functioning (DIF) and 2) to compare previous findings to those produced by employing the Rasch model. The example of DIF utilizes the question format: "Please indicate how important you feel the following student education goals are when conducting your classes", with 18 statements from a study related to perceptions of student outcomes by community college faculty. The response scale options consist of: essential, very important, somewhat important and not important.

Method

Data for this project were collected via a census survey of faculty during the summer 2005 pilot study of five community colleges throughout western VA and southern WV. In total, 41 faculty members returned valid responses. The survey instrument was a modified version of the 2004 Faculty Survey developed by the Higher Education Research Institute (HERI) at UCLA. Data were analyzed via the DIF method with Winsteps software (Version 3.58.1).

Partial Credit Model

For this project, the Partial Credit Model was utilized. Bond and Fox (2001) define the Partial Credit Model as "a Rasch model for polytomous data… which allows the number of ordered item categories and/or their threshold to vary from item to item" (p. 232). The model analyzes each question with respect to its response scale, as opposed to assuming equal response categories (and response scale) for all questions. This is the default model within Winsteps.

Differential Item Functioning (DIF)

In order to interpret DIF, several issues should be discussed, including how to ascertain if DIF is present. Throughout the literature, DIF is presented in multiple contexts, utilizing statistical and measurement models. According to Thissen, Steinberg, and Wainer (1993), as

cited in Roever (2005), "IRT techniques are the 'gold standard' of DIF detection" (p. 5). Here, DIF is detected through the Rasch model, a one-parameter IRT model. According to Zwick & Thayer (1996), *Average Observations* are the primary indicator of DIF, so these values were utilized in this study.

Rasch measurement assumes that test takers, or in this case the individual responding to the survey, with similar knowledge and abilities, or opinions, will perform alike regardless of sex, race, etc. DIF allows data to be examined by subgroup to detect differences between their perceptions on a given variable. Here, DIF is used to detect differences among perceptions about faculty paradigms. Using collapsed scales, moderate-strong *hard* scientist, slightly *hard* scientist, equally both, slightly *social/behavioral* scientist and moderate-strong *social/behavioral* scientist to define the subgroups, responses are then compared.

Results and Discussion

Most research suggests faculty from the "hard" sciences are concerned primarily with the intellectual growth of students and less concerned with non-cognitive factors such as emotional, social, and cultural growth (Smart, Feldman and Ethington, 2000). As expected, when compared to their *hard* science counterparts ("Faculty Class" = 1 in Table 1), individuals identified as the moderate-strong *social/behavioral* scientists ("Faculty Class" = 3 in Table 1) were more concerned with matters such as Creative Capacities, Social Skills, Communicating Orally, Cultural Appreciation and Cultural Knowledge. Even so, the DIF analysis, presented in Table 1, supports that faculty classified as moderate-strong *hard* scientists are also concerned (almost equally) with the Emotional Development and Spiritual Development of their students. In fact, the data indicate some of the moderate-strong *hard* scientists are slightly more concerned with Moral Character than their faculty counterparts.

Table 1. DIF Map

INPUT: 41 FACS, 18 PCTS MEASURED: 41 FACS, 18 PCTS, 4 CATS MINISTEP 3.58.1

FAC CLASS	OBSERVATIONS COUNT	AVERAGE	BASELINE EXPECT	MEASURE	DIF SCORE	DIF MEASURE	DIF SIZE	DIF S.E.	DIF t	PCT Number	Name
1	9	**.22**	.11	3.40	.11	2.59	-.81	.79	-1.03	1	THINK CRITICALLY
2	16	**.06**	.10	3.40	-.04	3.92	.52	1.05	.50	1	THINK CRITICALLY
3	14	**.07**	.16	3.40	-.09	4.31	.91	1.04	.87	1	THINK CRITICALLY
1	9	**.89**	.63	1.10	.26	.44	-.66	.51	-1.29	2	EMPLOYMENT AFTER COLLEGE
2	16	**.50**	.53	1.10	-.03	1.21	.11	.46	.24	2	EMPLOYMENT AFTER COLLEGE
3	14	**.64**	.78	1.10	-.13	1.45	.35	.44	.80	2	EMPLOYMENT AFTER COLLEGE
1	9	**1.11**	1.20	-.25	-.08	-.07	.18	.49	.37	3	FOUR YEAR COLLEGE PREP
2	16	**.88**	1.02	-.25	-.15	.11	.37	.40	.91	3	FOUR YEAR COLLEGE PREP
3	14	**1.43**	1.37	-.25	.06	-.38	-.13	.39	-.33	3	FOUR YEAR COLLEGE PREP
1	9	**1.11**	.89	.45	.23	-.07	-.51	.49	-1.04	4	MORAL CHARACTER
2	16	**.81**	.75	.45	.06	.28	-.17	.41	-.41	4	MORAL CHARACTER
3	14	**.93**	1.05	.45	-.12	.73	.28	.41	.69	4	MORAL CHARACTER
1	9	**1.78**	1.38	-.64	.40	-1.50	-.86	.49	-1.75	5	EMOTIONAL DEVELOPMENT
2	16	**1.00**	1.19	-.64	-.19	-.20	.44	.39	1.13	5	EMOTIONAL DEVELOPMENT
3	14	**1.57**	1.55	-.64	.02	-.69	-.05	.39	-.13	5	EMOTIONAL DEVELOPMENT
1	9	**1.33**	1.20	-.25	.14	-.55	-.30	.49	-.61	6	PERSONAL VALUES
2	16	**1.13**	1.02	-.25	.10	-.50	-.25	.38	-.64	6	PERSONAL VALUES
3	14	**1.29**	1.37	-.25	-.08	-.07	.18	.39	.46	6	PERSONAL VALUES
1	9	**1.44**	1.32	-.53	.12	-.79	-.26	.49	-.53	7	COMMUNITY SERVICE
2	16	**1.31**	1.14	-.53	.17	-.93	-.40	.38	-1.06	7	COMMUNITY SERVICE
3	14	**1.29**	1.50	-.53	-.21	-.07	.46	.39	1.16	7	COMMUNITY SERVICE
1	9	**1.22**	1.12	-.08	.10	-.31	-.23	.49	-.46	8	RESPONSIBLE CITIZENSHIP
2	16	**1.13**	.95	-.08	.17	-.50	-.42	.38	-1.08	8	RESPONSIBLE CITIZENSHIP
3	14	**1.14**	1.29	-.08	-.15	.24	.32	.40	.81	8	RESPONSIBLE CITIZENSHIP
1	9	**1.00**	1.48	-.86	-.48	.18	1.04	.50	2.07	9	CULTURAL KNOWLEDGE
2	16	**1.25**	1.28	-.86	-.03	-.79	.07	.38	.18	9	CULTURAL KNOWLEDGE
3	14	**1.86**	1.65	-.86	.21	-1.32	-.47	.40	-1.15	9	CULTURAL KNOWLEDGE
1	9	**.89**	1.38	-.64	-.49	.44	1.08	.51	2.10	10	CULTURAL APPRECIATION
2	16	**1.25**	1.19	-.64	.06	-.79	-.15	.38	-.39	10	CULTURAL APPRECIATION
3	14	**1.71**	1.55	-.64	.17	-1.00	-.36	.40	-.92	10	CULTURAL APPRECIATION
1	9	**.56**	.33	2.09	.23	1.31	-.79	.57	-1.37	11	MASTER DISCIPLINE KNOWLEDG
2	16	**.13**	.28	2.09	-.16	3.13	1.04	.77	1.35	11	MASTER DISCIPLINE KNOWLEDG
3	14	**.50**	.43	2.09	.07	1.87	-.22	.48	-.47	11	MASTER DISCIPLINE KNOWLEDG
1	9	**.89**	1.17	-.20	-.28	.44	.63	.51	1.24	12	CREATIVE CAPACITIES
2	16	**1.06**	1.00	-.20	.06	-.35	-.15	.39	-.40	12	CREATIVE CAPACITIES
3	14	**1.43**	1.34	-.20	.09	-.38	-.19	.39	-.47	12	CREATIVE CAPACITIES
1	9	**2.11**	1.86	-1.68	.25	-2.27	-.59	.52	-1.12	13	SPIRITUAL DEVELOPMENT
2	16	**1.44**	1.65	-1.68	-.21	-1.21	.47	.37	1.26	13	SPIRITUAL DEVELOPMENT
3	14	**2.07**	2.01	-1.68	.06	-1.82	-.14	.42	-.34	13	SPIRITUAL DEVELOPMENT
1	9	**1.11**	1.32	-.53	-.21	-.07	.46	.49	.93	14	SOCIAL SKILLS
2	16	**1.19**	1.14	-.53	.05	-.64	-.11	.38	-.30	14	SOCIAL SKILLS
3	14	**1.57**	1.50	-.53	.07	-.69	-.16	.39	-.41	14	SOCIAL SKILLS
1	9	**1.11**	1.20	-.25	-.08	-.07	.18	.49	.37	15	STRESS MANAGEMENT
2	16	**.94**	1.02	-.25	-.08	-.04	.21	.40	.52	15	STRESS MANAGEMENT
3	14	**1.50**	1.37	-.25	.13	-.54	-.28	.39	-.72	15	STRESS MANAGEMENT
1	9	**1.00**	.89	.45	.11	.18	-.26	.50	-.53	16	TIME MANAGEMENT
2	16	**.75**	.75	.45	.00	.45	.00	.42	.00	16	TIME MANAGEMENT
3	14	**.93**	1.05	.45	-.12	.73	.28	.41	.69	16	TIME MANAGEMENT
1	9	**2.22**	2.11	-2.27	.11	-2.55	-.29	.54	-.53	17	WRITE EFFECTIVELY
2	16	**2.00**	1.90	-2.27	.10	-2.50	-.24	.39	-.61	17	WRITE EFFECTIVELY
3	14	**2.07**	2.24	-2.27	-.17	-1.82	.44	.42	1.05	17	WRITE EFFECTIVELY
1	9	**.33**	.78	.70	-.45	2.07	1.37	.67	2.03	18	COMMUNICATE ORALLY
2	16	**.75**	.66	.70	.09	.45	-.25	.42	-.60	18	COMMUNICATE ORALLY
3	14	**1.14**	.94	.70	.20	.24	-.45	.40	-1.14	18	COMMUNICATE ORALLY

*Notes: 1. Missing data (N = 2) were removed from the table 2. *FAC Class: 1=Moderate-strong “hard” scientist; 2=Slight “hard” scientist, equally both, slight social/behavioral scientist; 3=Moderate-strong social behavioral scientist.

Educational Significance

Although much of the research has indicated the *hard* science faculty as less concerned with noncognitive issues, this example provides a counter-example to that theory. In general, understanding faculty perceptions of various outcomes is important as it relates to accreditation issues, given nearly all accrediting agencies require certain outcomes met. This research provides a baseline to unveil the intricacies involved as it relates to faculty perceptions of these outcomes.

Example 6 – *The Use of the Rasch Partial-Credit Rating Scale Model to Assess the Quality of a Web-Based Survey on Therapist Characteristics*

Jennifer Weber; University of Kentucky

Research on the impact of therapist characteristics has included the influence of client preferences for such qualities as sex, race and age. For example, a study on what inpatients found to be idealistic in a therapist revealed that almost two-thirds preferred a female therapist over a male therapist (Jonker, De Jong, de Weert-van Oene, & Gijs, 2000). Similarly, Grosenick and Hatmaker's (2000) work with clients in a treatment facility showed that respondents "rated the presence of a female staff as mattering *most* or *quite a bit*" (p. 281). Even so, findings of significant differences between preferences for a therapist based on sex, as well as based on ethnicity, age, religious affiliation, etc., are not consistent within the literatures.

Objective

This example assesses the psychometric qualities of an online survey constructed to measure client preferences for therapist characteristics. Rasch analysis was used to evaluate the reliability of the data collected, evaluate the survey's rating scale structure and evaluate the

individual items on the survey (Rasch, 1960). Results from the analysis were used to revise the survey for further study on preferences for therapist variables.

Method

The target sample for the online survey resulted in a homogeneous group of mostly Caucasian undergraduate and graduate students, faculty and administrators who subscribed to LISTSERVs for various programs housed by the College of Education at a Southeastern University. Fifty-eight valid responses were collected and utilized in the analysis.

Instrumentation

The web-survey was created and posted using Perseus SurveySolutions®/Express. The survey consisted of questions regarding the participants' demographics, experiences in and impressions of counseling, and preferences for certain therapist characteristics. Open-ended questions in the survey solicited unanticipated characteristic preferences. Survey items appropriate for Rasch analysis included 25 counselor characteristics that pertain to therapy-specific variables (e.g.: sympathetic, validating, trustworthiness, understanding, competent, good listener, collaborating, challenging, etc.). Responses to the items correspond to a Likert-type rating scale, where 1 = strongly disagree, 2 = disagree, 3 = agree and 4 = strongly agree.

Data Analysis

Data transformed into interval data through Rasch analysis were analyzed using fit statistics, hierarchical ratings, and probability curves yielded from Winsteps (Linacre, 2004, version 3.51), employing the polytomous partial-credit rating scale model. Each participant is represented by a person label which consists of coded information including demographics and opinions of counseling. Infit and outfit output tables allow for the evaluation of unidimensionality and specific misfit in relation to the model expectations. Rating scale structure

was evaluated using category probability curves to determine if mean measures increased as the categories stepped up the scale in the 'more' direction. Using a map of participants and items, the item hierarchy was examined to determine if the items fell in the hypothesized structure and spread evenly across the intended range against participants' level of willingness to endorse items. Furthermore, mean square values from misfitting items were reviewed using item statistic tables produced by Winsteps to shed light onto which items may not have accurately measured participant preferences.

Results and Discussion

Participants tended to agree or strongly agree with therapist preferences identified in the literature that were included in the survey. This is not surprising given the positive and therapeutic nature of the characteristics examined. Findings from Rasch analysis of the survey data provide evidence for revising the survey to better reflect participants' preferences. More specifically, Table 1 shows item statistics in a fit-order table produced by Winsteps. The range of acceptable mean square infit and outfit values is 0 to 1.3 for samples sizes less than 500 (Bond & Fox, 2001). Items with values falling outside the desired range are viewed as misfitting items and potentially problematic; therefore, in need of revision prior to further usage. Items with infit mean square values outside the expected range (items 3 and 8) indicate 'off-variable noise' while items with outfit mean square values outside the given range (items 3, 14, 12, 8, and 5) indicate the presence of unexpected outliers.

Table 1. *Item Statistics in Order of Misfit [abbreviated version]*

Item		INFIT		OUTFIT		
	Model S.E.	MNSQ	ZSTD	MNSQ	ZSTD	Item Description
3	.26	1.82	3.1	1.89	3.4	sympathetic
14	.28	1.25	1.1	1.56	2.2	comfortable with diversity
12	.27	1.27	1.2	1.45	1.9	validates my thoughts
8	.27	1.44	1.9	1.41	1.8	uses humor
5	.28	1.27	1.3	1.33	1.4	focused

Additionally, item statistics were used to evaluate individual items. Statistics indicated items 3, 12, 8, 14, 13, 5, 24 and 2 appear to be misfitting, having outfit mean square values that fall outside the given acceptable range, less than zero and/or greater than 1.3. This may imply the items are not tapping into the underlying construct or are being misunderstood or interpreted in a way other than the researcher intended.

The category probability curve allows for review of the rating scale structure. While it demonstrates that the categories are all defined, respondents tended to over utilize the ‘agree’ and ‘strongly agree’ options. Additionally, the item and person map in Figure 1 displays a hierarchy of characteristics preferences as rated by the participants and indicates that participants’ willingness to endorse the items is generally very high and the item endorsability is quite easy, as noted by the mean, m, of items and persons. The characteristic which participants rated as most preferable in a counselor was “*a good listener”*. The characteristics which participants rated as the least preferable in a counselor were “*sympathetic”*, *“validates my thoughts”*, *“uses humor”*, and *“comfortable talking about issues of diversity*”.

Figure 1. *Hierarchy Map of Persons and Items*

```
                                   Persons -MAP- Items
                                         <more>|<rare>
  6       15211 15211 15212 15313 25211 25712  +
                                               |
                                             T|
                                               |
                              15313 15612 25211  |
  5                                  15611  +
                 11312 14212 14311 15312 26611  |
                                               |
                                     10812 22211 S|
                                               |
  4                                  15311  +
                              15211 15212 25412  |
                        15211 15211 25312 25912  |
                                     15211 15312  |
                              10211 15311 15712 M|
  3 15211 15211 15311 15311 15312 15415 25512  +
                                     15711 25311  |
                                           15212  |
                              15311 15312 15312  |
                        15211 15211 15511 22111  |
  2                                  15211 S+
                 15211 15312 25211 25612 25912  |T
                                           15512  |
                                     11811 15311  |
                                           15211  |  is sym uses h valida
  1                                               +S is com
                                             T|  challe
                                               |  explai is ope works
                                               |  helps  identi speaks valida
                                               |  is com
  0                                               +M accura unders
                                               |  is gen
                                               |  believ encour
                                               |  I can
                                               |  gives  makes
 -1                                               +S respec
                                               |
                                               |  seems
                                               |
                                               |T I can
 -2                                               +
                                               |  is a g
                                               |
                                               |
                                               |
 -3                                               +
                                         <less>|<frequ>
```

Educational Significance

This study contributes both theoretical and methodologically to the fields of counseling and survey research. It contributes to the base of knowledge and provides further evidence as to client preferences for certain therapists' characteristics. Most importantly, results support necessary revisions for the use of the survey in further research.

Example 7 – ***Utilizing Principal Components within Rasch to Investigate Unidimensionality of the Psychological Sense of School Membership Scale***

R. Trent Haines; Louisiana State University

In order to measure a sense of belonging for students, Goodenow developed a scale that is designed to measure the psychological sense of school membership (PSSM) of adolescents (see Goodenow, 1993b for a full discussion of scale development). Goodenow reported reliability scores for suburban ($\alpha = 0.86$) and urban students ($\alpha = 0.88$). While these reliability scores have been used to suggest that the scale is unidimensional for these samples, it remains to be determined if the scale is truly unidimensional for students of various subgroups.

Objective

One of the primary concerns of Rasch measurement is the concept of unidimensionality. Unidimensionality refers to the concept that all items within a set of data should measure a single latent trait (Smith, 2004b). In constructing a test, Linacre (2004) says that "all items must be about the same thing, but then be as different as possible." Additionally, Bejar (1983) suggests that, although one particular latent trait is to be addressed by a set of items, there are a number of psychological processes that contribute to the latent trait and the examination thereof, which can be addressed by multiple items. In Goodenow's (1993b) school belonging measure, each question is designed to measure aspects of the individual's sense of school membership. Using a Likert-type response scale, individual responses are designed such that a higher score on an item indicates a higher sense of school membership in reference to the particular item.

The use of person and item fit statistics are important diagnostic tools. Here, the primary concern is with item outfit statistics as related to the unidimensionality of the PSSM. Outfit refers to the pattern of unexpected responses (Schumacker, 2004). Specifically, this inquiry is

directed at determining if PSSM items measure additional constructs beyond school belonging for African American male adolescents.

Rasch measurement provides two options for evaluating outfit: mean square and z-standardized fit statistics. Given the mean square fit statistic is more useful when determining if the data fit the Rasch model in a useful way (Linacre, 2003), mean square fit statistics are utilized in this example. This example answers the question: Is the PSSM scale unidimensional when used with low-income African American male adolescents?

Method

The sample for this study was a purposeful sample of 46 African American male students, grades 6 – 8, who attended a public last-chance alternative school (Raywid, 1994) in a small southeastern city. Each student completed Goodenow's (1993b) PSSM survey during the 2001 – 2004 school years as part of a program evaluation for a nonprofit organization.

Data Analysis

Data were entered for each respondent into Winsteps version 3.54 (Linacre, 2004). Items were coded such that 1 = not at all true and 5 = completely true. Negatively worded survey items were recoded such that a higher score equated to an increased sense of school belonging. As outlined in Linacre (1995), the following steps were used to assess unidimensionality. First, point-biserial correlations were examined. Negative correlations often suggest that items have been entered incorrectly, items have not been properly recoded or that items are ambiguous. Second, fit statistics were examined. Items having fit statistics larger than the accepted threshold tend to indicate unpredicted responses. Here, the accepted range for z-standardized fit statistics is the criteria of +/2 set by Linacre (2002). The third step, as explained by Smith (2004a), is that the researcher should use unrotated principle components analysis of the standardized residuals to

investigate response patterns that may indicate multidimensionality. Linacre (2004) suggests that eigenvalues greater than four indicate serious problems with unidimensionality.

Results and Discussion

Residual correlations reveal five negative correlations of residuals (see Table 1). As Linacre indicated, negative residual correlations can be an indicator of multidimensionality. Though the negative correlations exist, the strength of the correlations is low and do not clearly indicate multidimensionality. In Rasch analysis, residual correlations are interpreted in the same manner as correlations are in traditional statistical analyses.

The second step in the analysis was the examination of fit statistics. As discussed above, outfit statistics were given special attention (see Table 2). Only one item indicates strong misfit, being greater than 2, possibly indicative of multidimensionality. Items that fall within an outfit of .5 to 1.5 fit the Rasch model and support unidimensionality.

The third step in determining unidimensionality was the use of principle components analysis of residuals. The first principle component factor has an eigenvalue of 3.1. Though the size of the eigenvalue is stronger than that of 1.4, as found in simulation studies conducted by Smith (Smith, 2004a), it is not large enough to be considered a separate factor. Additionally, Linacre (2004) suggests that an eigenvalue less than 4 typically does not indicate a separate factor. An eigenvalue of 3.1, however, is large enough to suggest that some items may lean toward a separate factor or, as Linacre (2004) says, “the second ‘dimension’ in the data, is a conspicuous branch, but not a separate tree-trunk.”

Table 1. *Largest standardized residual correlations used to identify dependent items*

Item One	Item Two	Residual Correlation
People here know I can do good work.	I feel proud of belonging to this school.	-0.43
Other students in this school take my opinions seriously.	The teachers here respect me.	-0.42
Teachers here are not interested in people like me.*	I feel proud of belonging to this school.	-0.41
Sometimes I feel as if I don't belong here.*	The teachers here respect me.	-0.40
The teachers here respect me.	I wish I were in a different school.*	-0.38

Note: * Items were reverse coded prior to analysis.

Table 2. *Mean logit scores and fit indices for individual items of the PSSM scale*

Item	Mean	Error	Outfit MSE
Other students in this school take my opinions seriously.	54.62	1.42	2.27
I wish I were in a different school.*	57.15	1.46	1.52
I feel different from most other students here.*	50.37	1.47	1.52
Sometimes I feel as if I don't belong here.*	49.67	1.45	1.44
There's at least one teacher or other adult in this school I can talk to if I have a problem.	43.98	1.57	1.35
People at this school are friendly to me.	47.66	1.47	1.26
It is hard for people like me to be accepted here.*	44.37	1.58	1.18
Most teachers at this school are interested in me.	43.98	1.57	1.14
I am included in lots of activities at this school.	55.63	1.43	1.05
I can really be myself at this school.	50.79	1.42	1.03
I feel like a real part of this school.	54.42	1.42	0.92
The teachers here respect me.	49.36	1.44	0.91
I feel proud of belonging to this school.	58.81	1.49	0.88
I am treated with as much respect as other students.	49.36	1.44	0.71
People here know I can do good work.	44.95	1.44	0.66
People here notice when I'm good at something.	46.78	1.49	0.66
Other students here like me the way I am.	46.34	1.50	0.63
Teachers here are not interested in people like me.*	50.18	1.43	0.48

Note: * Items were reverse coded prior to analysis.

The analyses of the PSSM scale (Goodenow, 1993b) indicates that the scale holds the property of unidimensionality here. In other words, each item in the scale contributes to the overall measurement of school belonging. These analyses affirm Goodenow's analysis as a measure of school belonging. Though the PSSM appears to be unidimensional, there are some areas that may warrant further study. The negative, albeit small, residual correlations between some items may indicate a subdimension in the scale. Three of the five pairs of negative correlations involve a negatively worded item. One reason for the negative correlation could be that such negatively worded items are interpreted differently than positively worded items as several researchers have found that such items are often misinterpreted by the respondent and may not measure the construct in the same way as positively worded items.

There was only one item with an outfit score that was above the acceptable threshold: "Other students in this school take my opinions seriously." Since this item's outfit is only slightly above the threshold of 2, it is likely that the outfit is merely "noise" and does not reflect multidimensionality in the scale. However, it could indicate that the students may view their peers somewhat differently than they view their teachers or the school as a whole. Though there were some anomalies in the analyses, the results indicate that, on the whole, the PSSM is a unidimensional instrument that captures the essence of school belonging for this sample, African American students who attended an alternative school.

Educational Significance

In psychological measures, and in general, it is often assumed that a rating scale holds together as theoretically purported. One of the primary concerns of Rasch measurement is the concept of unidimensionality. Here, an empirical process is presented to assess the underlying construct being measured. Such techniques could be applied to many existing measures.

Conclusion

This collaboration of student-lead research projects utilizes the Rasch model employing Winsteps software. It provides an overview of applications connecting to such areas as higher education, educational psychology and counseling. Here theoretical frameworks, methodological processes and results for rating scale and assessment formats are demonstrated, supplying a clear demonstration of the utility of the model. The concepts are relevant for k-12 practitioners, college and university students, faculty and those interested in methodological practices. If an individual is making inferences based on total scores, either for a questionnaire or an assessment, a Rasch analysis is a sound alternative to traditional analyses. This paper presents the beginning tools to explore this option.

References

Allen, B. A., & Boykin, A. W. (1991). The influence of contextual factors on Afro-American and Euro-American children's performance: Effects of movement opportunity and music. *International Journal of Psychology, 26*(3), 373-387.

Allen, B. A., & Boykin, A. W. (1992). African-American children and the educational process: Alleviating cultural discontinuity through prescriptive pedagogy. *School Psychology Review, 21*, 586-597.

American Association for Employment in Education (2004). *Educator Supply and Demand*. Columbus, OH: American Association for Employment in Education.

Anderman, L. H. (1999). Classroom goal orientation, school belonging and social goals as predictors of students' positive and negative affect following the transition to middle school. *Journal of Research and Development in Education, 32*, 89-103.

Andrich, D. (2004). Controversy and the Rasch model: A characteristic of incompatible paradigms? In E. V. Smith, Jr. and R. M. Smith *Introduction to Rasch Measurement* (pp.143-166) Maple Grove, MN: JAM Press.

Andrich, D. & Luo, G. (2003). Conditional Pairwise Estimation in the Rasch Model for Ordered Response Categories using Principal Components. *Journal of Applied Measurement, 4*(3), 205-221.

Becker, B. E., & Luthar, S. S. (2002). Social-emotional factors affecting achievement outcomes among disadvantaged students: Closing the achievement gap. *Educational Psychologist, 37*(4), 197-214.

Bejar, I. I. (1983). *Achievement Testing: Recent Advances*. Beverly Hills, CA: Sage Publications.

Bethea, P. D. (1995). African-American women and the male-female relationship dilemma: A counseling perspective. *Journal of Multicultural Counseling &Development, 23*(2), 87-95.

Beutler, L. E., Machado, P. P., & Neufeldt, S. (1994). Therapist variables. In A. E. Bergin and S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 229-269). New York: Wiley.

Bond, T., & Fox, C. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Booker, K. C. (2006). School belonging and the African American adolescent: What do we know and where should we go? *The High School Journal, 89*, 1-7.

Broman, C. L. (1993). Race differences in marital well-being. *Journal of Marriage &the Family, 55*(3), 724-732.

Castillo, J. (2002). A note on the conception of tacit knowledge. *Journal of Management Inquiry, 11*(1), 46-57.

Colonia-Wilner, R. (1999). Investing in practical intelligence: Ageing and cognitive efficiency among executives. *International Journal of Behavioral Development, 23,* 591-614.

David, A. B., & Erickson, C. A. (1990). Ethnicity and the therapist's use of self. *Family Therapy, 17*(3), 211-216.

Edwards, W. R., & Schleicher, D. J. (2004). On selecting psychology graduate students: Validity evidence for a test of tacit knowledge. *Journal of Educational Psychology, 96,* 592-602.

Fox, C. (1999). An introduction to the partial credit model for developing nursing assessments. *Journal of Nursing Education, 34*(8), 340-6.

Gay, G. (2000). *Culturally Responsive Teaching*. New York, NY: Teachers College Press.

Goodenow, C. (1993a). Classroom belonging among early adolescent students: Relationships to motivation and achievement. *Journal of Early Adolescence, 13*, 21-43.

Goodenow, C. (1993b). The psychological sense of school membership among adolescents: Scale development and educational correlates. *Psychology in the Schools, 30*, 79-90.

Grosenick, J. K., & Hatmaker, C. M. (2000). Perceptions of staff attributes in substance abuse treatment. *Journal of Substance Abuse Treatment, 19*, 273-284.

Guttman, L. (1950). The principal components of scale analysis. In S.A. Stouffer, L. Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star and J.A. Clausen (Eds.), *Measurement and Prediction,* pp.312-361. New York: Wiley.

Hedlund, J., Forsythe, G. B., Horvath, J. A., Williams, W. M., Snook, S., & Sternberg, R. J. (2003). Identifying and assessing tacit knowledge: understanding the practical intelligence of military leaders. *The Leadership Quarterly, 14,* 117-140.

Jensen, A. R. (1993). Test validity: g versus "tacit knowledge". *Current Directions in Psychological Science, 2*(1), 9-10.

Jonker, J., De Jong, C. A., de Weert-van Oene, G. H., & Gijs, L. (2000). Gender-role stereotypes and interpersonal behavior: How addicted inpatients view their ideal male and female therapist. *Journal of Substance Abuse Treatment, 19*, 307-312.

Keijsers, G. P. J., Schaap, C. P. D. R., & Hoogduin, C. A. L. (2000). The impact of interpersonal patient and therapist behaviors on outcome in cognitive-behavioral therapy: A review of empirical studies. *Behavior Modification, 24*(2), 264-297.

Kiecolt, K. J. & Fossett, M. A. (1997). Mate availability and marriage among African Americans. In R.J. Taylor (Ed.), *African American Research Perspectives.* Ann Arbor: University of Michigan Press.

King, A. E. (1999). African American females' attitudes toward marriage: An exploratory study. *Journal of Black Studies, 29*(3), 416-437.

Kinnon, J. B. (2003). The shocking state of Black marriage: Experts say many will never get married. *Ebony.* Retrieved on November 1, 2004. http://www.findarticles.com/p/articles/mi_m1077/is_1_59/ai_110361377/print

Kochman, T. (1981). *Black and White Styles in Conflict*. Chicago, IL: The University of Chicago Press.

Lafferty, P., Beutler, L. E., & Crago M. (1989). Differences between more and less effective psychotherapists: A study of select therapist variables. *Journal of Consulting and Clinical Psychology, 57*(1), 76-80.

Lawson, E. J. & Sharpe, T.L. (2000). Black men and divorce: Implications for culturally competent practice. *Minority Health Today*. Retrieved on September 1, 2005, from http://findarticles.com/p/articles/mi_m0HKU/is_5_1/ai_66918338/pg_7.

Linacre, J. M. Winsteps Rasch measurement software version. Available at www.winsteps.com

Linacre, J. M. (1995). Prioritizing misfit indicators. *Rasch Measurement Transactions, 9*, 422-423.

Linacre, J. M. (1999). A *User's Guide to Facets Rasch Measurement Computer Program*. Chicago, IL: MESA Press.

Linacre, J. M. (2002a). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*, 878.

Linacre, J. M. (2002b). Facets, factors, elements and levels [Electronic version]. *Rasch Measurement Transactions, 16*(2), 880.

Linacre, J. M. (2003). Size vs. significance: Standardized chi-square fit statistics. *Rasch Measurement Transactions, 17*, 918.

Linacre, J. M. (2004a). *A User's Guide to Winsteps Rasch-Model Computer Programs.* Chicago, IL: MESA Press.

Linacre, J. M. (2004b). Optimizing rating scale category effectiveness. In E. V. Smith, Jr. and R. M. Smith *Introduction to Rasch Measurement* (pp.258-278) Maple Grove, MN: JAM Press.

Luborsky, L., Crits-Christoph, P., & McLellan, A. (1986). Do therapists vary much in their s success? Findings from our outcome studies. *American Journal of Orthopsychiatry, 56*, 501-512.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149-174.

Midgely, C., Feldlaufer, H., & Eccles, J. S. (1989). Student/teacher relations and attitudes towards mathematics before and after the transition to junior high school. *Child Development, 90*, 981-992.

Moos, R. H. (1979). *Evaluating Educational Environments*. San Francisco, CA: Jossey-Bass.

Nardi, P. M. (2003). *Doing Survey Research: A Guide to Quantitative Methods*. Boston: Allyn & Bacon.

Najavits, L. M., & Weiss, R. D. (1994). Variations in therapist effectiveness in the treatment of patients with substance use disorders: An empirical review. *Addiction, 89*, 679-688.

Nestor-Baker, N., & Hoy, W. K. (2001). Tacit knowledge of school superintendents: Its nature, meaning, and content. *Education Administration Quarterly, 37*(1), 86-129.

Phelan, P., Davidson, A. L., & Cao, H. T. (1991). Students' multiple worlds: Negotiating the boundaries of family, peer, and school cultures. *Anthropology and Education Quarterly, 22*, 224-250.

Phelan, P., Davidson, A. L., & Yu, H. C. (1998). *Adolescent Worlds: Negotiating Family, Peers, and School*. New York, NY: Teachers College Press.

Rasch, G. (1960). *Probabilistic models for some intelligence and achievement tests*. Copenhagen: Danish Institute for Educational Research (Expanded edition, 1980. Chicago: University of Chicago Press).

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests.* Chicago: The University of Chicago Press (original work published in 1960).

Raywid, M. A. (1994). Alternative schools: The state of the art. *Educational Leadership*, 26-31.

Roever, C. (2005). *"That's not fair!" Fairness, bias, and differential item functioning in language testing*. Second Language Studies Brownbag, November 15, 2005. University of Hawaii.

Schmidt, F. L., & Hunter, J. E. (1993). Tacit knowledge, practical intelligence, general mental ability, and job knowledge. *Current Directions in Psychological Science, 2*(1), 8-9.

Schumacker, R. E. (2004). Rasch Measurement: The Dichotomous Model. In E. V. Smith, Jr. & R. M. Smith (Eds.), *Introduction to Rasch Measurement: Theory, Models and Applications*. Maple Grove, MN: JAM Press.

Smart, J. C., Feldman, K. A., & Ethington, C. A. (2000). *Academic Disciplines: Holland's Theory and the Study of College Students and Faculty*. Nashville, TN: Vanderbilt University Press.

Smith, E. V., Jr. (2004a). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal components analysis of residuals. In E. V. Smith, Jr. & R. M. Smith (Eds.), *Introduction to Rasch Measurement: Theory, Models and Applications*. Maple Grove, MN: JAM Press.

Smith, E. V., Jr. (2004b). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. In E. V. Smith, Jr. & R. M. Smith (Eds.), *Introduction to Rasch Measurement: Theory, Models and Applications*. Maple Grove, MN: JAM Press.

Smith, E. V., Jr., & Smith, R. M. (2004). *Introduction to Rasch Measurement: Theory, Models, and Applications*. Maple Grove: JAM Press.

Smith, R. M. (1996). Polytomous mean-square statistics. *Rasch Measurement Transactions, 6*, 516-517.

Somech, A., & Bogler, R. (1999). Tacit knowledge in academia: Its effects on student learning and achievement. *The Journal of Psychology, 133,* 605-616.

Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J. A., Wagner, R. K., Williams, W. M., Snooks, S. A. & Grigorenko, E. L. (2000). *Practical intelligence in everyday life.* New York: Cambridge University Press.

Sternberg, R. J., & Wagner, R. K. (1993). The g-ocentric view of intelligence and job performance is wrong. *Current Directions in Psychological Science, 2*(1), 1-5.

Taylor, P. L. (1998). Attitudes toward marriage among African American singles: A test of four perspectives. Los Angeles: University of California Press.

Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P.W. Holland, & H.

Wainer (Eds.), *Differential item functioning* (pp. 67-113). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tschannen-Moran, M., & Nestor-Baker, N. (2004). The tacit knowledge of highly productive scholars in education. *Teachers College Record, 106,* 1484-1511.

Wagner, R. K. (1987). Tacit knowledge in everyday intelligent behavior. *Journal of Personality and Social Psychology, 52,* 1236-1247.

Weiner, B. (1990). The history of motivation research in education. *Journal of Educational Psychology, 82*, 616-622.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA Press.

Wright, B. (1997). Fundamental measurement for outcome evaluation [Electronic version]. *Physical Medicine and Rehabilitation: State Of The Art Reviews, 11*(2), 261-288. Available at: http://www.rasch.org/memo66.htm

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA Press.

Wright, B. D. & Stone, M. H. (1979). *Best test design.* Chicago: MESA Press.

Zwick R., & Thayer, D. T. (1996). Evaluating the magnitude of Differential Item Functioning in polytomous items. *Journal of Educational Statistics 21(*3), 187-201.

Appendix: Sample Winsteps Syntax

Example 1: Tacit Knowledge

```
TITLE="TKIS-Round II-Organizational"      page heading
NAME1=1                    ; Person-id starts in column 1 IDCODE/OFFICECODE/YRSSUP/EDLVL
ITEM1=10                   ; Item responses start in column 10
NI=15                      ; There are 15 items
IDELETE=4d,5d
CODES=1234567                      ; Scores entered as 1-7
KEY1=36533311*6524*4; Fully correct-3
KEY2=**6*4422**635*5; Fully- 3
KEY3=25422233*5413*3; 2pts
KEY4=477455***7746*6; 2pts
KEY5=14311144*4352*2; 1pt; if no matching response, use a character not in CODES=, e.g., *
KEY6=5**566******7*7; 1 pt
KEYSCR=332211              ; KEY1, KEY2- fully correct (3 points each),
                           ; KEY3, KEY4- partially correct (2 points each)
                    ; KEY5, KEY6- partially correct (1 point each)
ISGROUPS=0                 ; Each item is its own grouping, i.e., the Partial Credit model
MODELS=R                   ; Each item has its own Andrich rating scale
STKEEP=Y                   ; Keep all intermediate categories in analysis, even if never observed
&END
CVS1Q1
CVS1Q2
CVS1Q3
CVS1Q4
CVS1Q5
CVS4Q1
CVS4Q2
CVS4Q3
CVS4Q4
CVS4Q5
CVS5Q1
CVS5Q2
CVS5Q3
CVS5Q4
CVS5Q5
END NAMES
101010003523133312561126
102010003776176523671577
103010003545255524262367
104010003536735312767325
105010003625155212652246
106010003535255434655345
107030103355255322653252
108030604727123322663225
109030403754132411771335
11003160557*2652136*****
.
.
.
137030203433355224654454
138031205546145514751427
139030103445236322565445
```

Example 2: Sins Created Equal?

```
; This is file "GILBERTTEST.txt"
; The control file name is e:\GILBERTTEST.txt
 TITLE= 'RAS DATA'
 NI= 33  ; 33 items
 ITEM1= 13   ; responses start in column 13 of the data
 NAME1= 1
 NAMLEN = 12
 ITEM= ATT ; items are called "attitudes"
 PERSON= PART ; persons are called "participants"
 XWIDE=2 ;data codes are two characters wide
 CODES= 0506070809101112131415; valid response codes (ratings) are 05,06,07,08,09,10,11,12,13,14,15
 CLFILE=*
 05 -5
 06 -4
 07 -3
 08 -2
 09 -1
 10  0
 11  1
 12  2
 13  3
 14  4
 15  5
 *          ; "*" means the end of a list
 &END       ; this ends the control specifications
; These are brief descriptions of the 32 items
10
11
12
.
.
.
41
42
END NAMES  ;this follows the item names: - the data follow:
01BMM131203215050615050507151507090615150911051005061009050905050805060909051 5
02BMM111404115050515050507151505060515150506050805050809050905050905050808051 5
03BMM141202215050814050607151506090615150607060905070806060805050805060807061 5
04BDF122404214050515050705151506050514150606060805060508060905050805050607051 5
05BMM121305115050615050506151505050515150505050705050508050705050705050606051 5
06BSF132304115050714050605141506091015150605050906060909060906050905050908061 5
07BMF141303214050813050705141506080614150508110905070710061011050805070907091 5
08BSM112101312051013050510131305150513140811051107101111051105050706061108061 5
.
.
.
21BSM131403213050712050507121405070514150506100805060610050705050705050705101 3
22BDF112304214050712050705131307080713140607100906070607060910050705061006101 4
23BMM112303115050715050505141505050514150505100605060506050710050605060505101 4
24BWF111605115050615050505151505050515150505050505050506050605050505050505051 5
25BSM112102212050712050507111205100511140607080906060909050906050905071007051 2
26BMM112403115050615050505131405060514150606101015150607050810050605060707071 5
27BSF131301115050515050505151505090515150505050605050606050905050605050905051 5
28BMF161403215050712051006141506060614140706110605060607060711050805060706101 3
29BMM131403213051011050506111308091014150606150805060710060810050705050606101 0
30BSF151203212051111051105121311111011140910151010140910101015050805101006151 5
```

Example 3: Examining Specifications (AAEE Survey)

```
TITLE= 'AAEE 2005 TEACHING FIELDS' ; ';' indicates a comment
 NI=    64            ; 64 items
 ITEM1= 5   ; responses start in column 5 of the data
 NAME1= 1  ; person-label starts in column 1 of the data
 NAMLEN = 4 ; Length of name variable is 4
 ITEM=  DEMAND ; items are called "perceived demands"
 PERSON= INST ; persons are called "teacher education institutes"
 CODES= 12345 ; valid response codes (ratings) are 1, 2, 3, 4, 5
 CLFILE=*    ; label the response categories
 1 Considerable Surplus   ; names of the response categories
 2 Some Surplus
 3 Balanced
 4 Some Shortage
 5 Considerable Shortage
 *          ; "*" means the end of a list
 &END       ; this ends the control specifications
; These are brief descriptions of the 25 items
AG
ARTVIS
BILING
BUS
CS
DANCE
DRTRF
PREK
K
PRIM
INTER
MID
ENGL
ESL
HEALTH
HOMEEC
JOURN
CLASSIC
.
.
.
PE
READ
BIO
CHEM
HUMRES
SUPER
AUDIOL
COUNSEL
PHYTHP
SNURSE
SPSYCH
SSOCW
SPPATH
END NAMES  ;this follows the item names: - the data follow:
```

Example 4: Black student's perceptions of marriage

```
; This is file "blackmalesurvey.txt"
; The control file name is e:\blackmalesurvey.txt
 TITLE= 'Black Male Survey'
 NI= 20  ; 20 items in Black Male Survey
 ITEM1= 8   ; responses start in column 8 of the data
 NAME1= 1
 NAMLEN=2
 ITEM= PER ; items are called "perceptions"
 PERSON= MALE ; persons are called "males"
 CODES= 1234; valid response codes (ratings) are 1, 2, 3, 4,
 CLFILE=*
 1 Strdis
 2 Dis
 3 Agr
 4 StrAgr
 *          ; "*" means the end of a list
 &END       ; this ends the control specifications
; These are brief descriptions of the 1-20 items
1benefits
2maintain
3head
4goodbw
5sexinc
6slavery
7faithful
8edattrac
9divorce
10fear
11monogamy
12triflin
13infedility
14lesstime
15betterjob
16goodjob
17want
18lockdown
19sexdec
20sucinsoc
END NAMES  ;this follows the item names: - the data follow:
  11133131242312423433223333
  21111131323333122222231233
  31131131233222222222222322
  411111313322 4223223232332
  51111132323423223332222223
  61111121322324223323322333
  71111131433213223232332233
  81121142432224224323222433
  91111122333214323333331333
.
.
.
2316221324213242222233241443
2411331141323341224133423 2
2511314423323342111233223 24
```

Example 5: DIF Faculty Perceptions

```
; This is file MWERA DIF Study.txt
TITLE= 'Question11'
NI= 18 ; 18 items
ITEM1= 2 ; responses start in column 2 of the data
NAME1= 1 ; person-label starts in column 1 of the data
ITEM= PCT ; items are called "perceptions"
PERSON= FAC ; persons are called "faculty"
CODES= 1234 ; valid response codes (ratings) are 1, 2, 3, 4
CLFILE= * ; label the response categories
1 Essential ; names of the response categories
2 Very Important
3 Somewhat Important
4 Not Important
* ; "*" means the end of a list
&END ; this ends the control specifications
; These are brief descriptions of the 18 items
THINK CRITICALLY
EMPLOYMENT AFTER COLLEGE
FOUR YEAR COLLEGE PREP
MORAL CHARACTER
EMOTIONAL DEVELOPMENT
PERSONAL VALUES
COMMUNITY SERVICE
RESPONSIBLE CITIZENSHIP
.
.
.
SOCIAL SKILLS
STRESS MANAGEMENT
TIME MANAGEMENT
WRITE EFFECTIVELY
COMMUNICATE ORALLY
END NAMES ;this follows the item names: - the data follow:
21132222233134332421
31232332233123222322
31222334444234342433
21111122222112211211
31212432244134432433
21313333333133333311
11213332211114221312
21121333322133222431
21131221233122211331
21121112132121111211
21221222122122221322
.
.
.

31221321121222121122
32233333333223333333
31142323233223332322
11122333332113233311
31222322233223332433
21122222233133321331
```

Example 6: Therapist Characteristics

```
; This file is pilot study survey rasch data.txt
TITLE="Therapist Characteristics question 12"
NI=40           ; There are 40 items
ITEM1=13        ; Item responses start in column 13
NAME1=1                 ; Start of person label
ITEM=PFR        ; Items are called "preferences"
PERSON=PAR      ; Persons are called "participants"
CODES=1234      ; Valid response codes (ratings) are 1, 2, 3, and 4
CLFILE=*        ; label the response categories
1 Strongly Disagree     ;names of the response categories
2 Disagree
3 Agree
4 Strongly Agree
5 No Opinion
*       ; "*" means end of a list
&END    ; this ends the control specifications
; descriptions of the items
is a good listener
understands me
is sympathetic
encourages me
.
.
.
helps me learn from my mistakes
is nonjudgmental
is professional
END NAMES       ; this ends the item descriptions: - the data follows:
2591218140  43333 33433  3343434433333433333334 3344
1521214133        4 444444  44 4  44 4  4   44  44 44
2521116154  4444344444444344444444444344444444444444
2541214146  4324434444444444444433444444444444433444
1521216243  3433343 34443  4434344433 43 34333 33344
1551217246  4324433343322333 44432 4433333 3343 3334
2531216244  4444434343433 4433444444444444444444444
1531217346  44 344434433323343 43433444 23 333333344
1531217154  4234433234334443333433333433342334433344
1521114143  3433344 4433344333344333344333333333343
2561218140  3333333343433234333443434333233333333334
1531116154  4334343333343333333333333333344333332343
1521214340  4444333444444344434434444444344443344344
.
.
.
1521112134  4343343444433434333333334333344333333343
2221116134  4424444344434444444444444444444444444444
1521114136  433 3332444  334332333233 33232443323343
2211114126  4333432444333344343433343343443443334443
1531317155  4444444444444444444444444444444444444444
1521114144  4444443344433334444443444444444444434444
1081217144  4444433344444444444444344444444444344444
1131217145  4444444444423444444444444444444444444444
1551116156  3334444444433333333433433443344444333444
1521114146  444433 323433  43334 33343 3 33333333344
```

Example 7: Principle Components PSSM

```
&INST
  TITLE = "PSSM Analysis"
 PERSON = person ; persons are ...
   ITEM = item ; items are ...
  ITEM1 = 6 ; column of response to first item in data record
     NI = 18 ; number of items
  NAME1 = 1 ; column of first character of person label
NAMELEN = 5 ; length of person label
  XWIDE = 1 ; number of columns per item response
  CODES = 12345 ; valid codes in data file
 CLFILE = * ; category label file for category naming
1+1 "NOT AT ALL TRUE"
1+2 "SOMEWHAT TRUE"
1+3 "MORE TRUE"
1+4 "MOSTLY TRUE"
1+5 "COMPLETELY TRUE"
* ; end of CLFILE=* list
 UIMEAN = 50 ; item mean for local origin
 USCALE = 10 ; user scaling for logits
 UDECIM = 2 ; reported decimal places for user scaling

&END
REAL PART
PEOPLE NOTICE
ACCEPTED
.
.
.
RESPECT
FEEL DIFFERENT
BE MYSELF
TEACHER RESPECT
GOOD WORK
DIFFERENT SCHOOL
PROUD
LIKE ME
END LABELS
DAVID444444342344444333
MALIK345334433343113124
ANWAR334232433434334224
BILLY323335544133534434
TONIO343442535314213512
JAMAR233224413134412112
DAMEN334432343343434234
RUBEN243234334234433122
ELIOT133225134112112113
.
.
.
ROGER553334555243343355
DEVON245324334223323443
BRISE555551524553255122
ARION554455555355554555
CALEB235124335233333423
```