Oct 7th, 1:45 PM - 3:00 PM

Improving Measurement of Ambiguity Tolerance Among Teacher Candidates

Kent Allan Rittschof
Georgia Southern University, kent_r@georgiasouthern.edu

Improving Measurement of Ambiguity-Tolerance Among Teacher Candidates

KENT A. RITTSCHOF

Georgia Southern University

ABSTRACT

The process of learning often requires dealing with the ambiguity of uncertain interpretations. A learner's tolerance for ambiguity involves the degree of willingness to accept or adapt to unfamiliar, unpredictable, or uncertain situations and ideas. This study examined the measurement of ambiguity tolerance (AT) among teacher candidates. Pre-service teachers (n = 114) attending a medium size university in the southeastern United States were administered McLain's (1993) Multiple Stimulus Types Ambiguity Tolerance (MSTAT-I) scale. Analyses were first conducted on item responses from the MSTAT-I, then on item responses from the MSTAT-II (McLain, 2009), an instrument comprised of a subset of 13 particular items from the original 22-item instrument. Rasch-model measures and diagnostic analyses were compared and illustrated graphically for the two versions of the AT instrument and then for two shorter versions. Findings indicated validity support for the MSTAT-II instrument, measurement improvement to reducing the number of Likert scale categories to 5, and further measurement benefits of an alternative 9-item AT instrument. The distribution of measured AT among participants was discussed with regard to individual differences among teacher candidates and the potential application of AT measurement as an intellectual disposition among educators.

Uncertainty, Teaching, and Learning. Throughout the learning process, when interpretations of meaning are uncertain students can experience ambiguity. Both teaching and self-directed learning typically involve the reduction of ambiguity relative to the information being learned. Yet, for a student to reduce ambiguity, or for a teacher to help students in that regard, the student must ultimately confront the reality that complete elimination of all possible uncertainty is not likely to occur. Thus, both teacher and student may find ways to tolerate some degree of ambiguity while reducing other ambiguities during the teaching-learning process. Hence, one of the personal characteristics that are of particular relevance to teachers and pupils is their *ambiguity tolerance* (AT).

AT involves a person's willingness to adapt to encounters with ambiguous situations or ideas (Jonassen & Grabowski, 1993). The importance of AT to teaching relates to the earliest uses of the construct when Frenkel-Brunswik (1949) suggested that people intolerant of ambiguity tend to arrive at premature closure, tend to think concretely, and tend to seek simplistic solutions. Though teachers typically simplify material to help students understand concepts at their current cognitive levels, tendencies in teachers to oversimplify issues and solutions to problems would arguably not serve students well in our increasingly complex world. Supporting this concern Peters and Amburgey (1982) found that those teachers who had a higher AT tended to use higher cognitive levels of verbal responses in their teaching. Also, Ream (1984) showed that group discussions with students involving ambiguous situations or personal experience examples increased the students' tolerance for ambiguity. Furthermore, investigations on AT among educators and learners has supported learning that involves complex problems, novel transfer using new examples, divergent learning tasks, and brainstorming (Jonassen & Grabowski, 1993). With these findings in mind, it should not be a surprise that AT has correlated positively ($r = .59$) with a constructivist teaching orientation (Gottleib, 2006).

Educators encounter ambiguous situations every day in dealing with student needs and new information. The ways that educators deal with these ambiguous situations have many implications for both the teacher and the students. By understanding their own AT, teachers can consider their many possible reactions to unpredictable situations, and how such reactions and associated choices might differentially affect their pupils. The relevance of a teacher's AT to teaching includes the ways in which associated thought and behavior are influenced by a teachers' existing schema for the classroom environment.

Through much effort by researchers, teachers, and students, instruments used to measure AT have evolved and improved over the course of the past few decades (Budner, 1962; McLain, 2009; Ward, 1994). One early instrument was an attempt to measure AT based upon the idea that perceived ambiguity arises from stimuli that are *complex, unfamiliar*, and *insoluble* (Budner, 1962). A different and improved instrument, *Multiple Stimulus Types Ambiguity Tolerance* (MSTAT-I), was developed three decades later and included items reflecting these same three stimulus types as well as *uncertain* stimuli and ambiguous stimuli *in general* (McLain, 1993). The MSTAT-I was further refined as the MSTAT-II using the psychometrically strongest 13 of the 22 items based on empirical evidence supporting the theoretical definition of the AT construct (McLain, 2009). The reduction of items to 13 was also intended to further reduce the cognitive burden on respondents. Items on both the MSTAT-I and MSTST-II were designed to be broadly understandable, value neutral, and context independent to help allow the instruments to be usable in many settings. Additionally, McClain (2009) provided evidence that the MSTAT-II did not encourage socially desirable responding responses, as opposed to truthful responses. Together these instrument characteristics may help improve the measurement validity in future studies, in comparison with characteristics of past instruments used. The current study investigates the quality of both the MSTAT-I and MSTAT-II toward use in future studies involving educators. Although factor analytic research (Lauriola, Foschi, Mosca, Weller, 2016) has identified multiple AT related attitude factors among instruments that may represent meaningful AT dimensions, here we use a Rasch modeling approach in order to employ modern item response theory (IRT) procedures and corresponding diagnostics to address our research questions using our modest sized sample of participants.

Research Objectives and Hypotheses. To determine whether versions of a selected instrument are well suited to investigating ambiguity tolerance as a general dimension among pre-service educators the following objectives were pursued.

1. Examine and compare measurement characteristics of the MSTAT-I and MSTAT-II instruments with teacher candidates.
2. Examine whether empirical data supports the use of the MSTAT-I and MSTAT-II with teacher candidates.
3. Determine whether Rasch model measures and diagnostics identify any need for revisions within either instrument version.
4. Explore and examine possible improved versions of the instrument for teacher candidates.

Method

Participants. Pre-service teachers (n = 114) attending a medium size university in the southeastern United States volunteered for the study. Participants were primarily female (85%) between the ages of 19 to 22 years, though a few participants were older, and age data was not collected.

Instruments and Items. Individual differences were measured using a 7-step Likert scale of the MSTAT-I instrument. The MSTAT-I is McLain's (1993) 22-item AT instrument that yielded an alpha reliability of .86 within his discriminant validity investigation. In addition to reliability considerations, the MSTAT-I was chosen for this because items within this instrument were developed to specifically avoid being highly suggestive of socially desirable responses. Figure 1 identifies items that make up the MSTAT-I instrument.

Procedure: Within a classroom environment, participants were asked by an experimenter to complete a paper and pencil MSTAT-I survey as follows: Participants were asked to rate their agreement level on the scale. The survey required approximately 15 minutes to complete. Participants were instructed to answer questions honestly, and to skip items that they did not want to respond to.

Results

Rasch rating-scale model (Andrich, 1978) analyses were conducted across six total calibrations. Four instrument versions stemming from the original 22 item MSTAT-I scale were analyzed within the six calibrations to construct measures and diagnostics from the ordinal raw scores (Linacre, 2016a).

Within the rating-scale model equation below, Likert scale steps, or categories, often numbering between four and seven on instruments, are represented by k.

$$\ln\left(\frac{P_{nik}}{P_{ni(k-1)}}\right) = B_n - D_i - F_k \qquad \textit{Rasch Rating Scale Equation}$$

The threshold difficulty level in the rating scale formula is represented as F_k, which can be calculated across a set of items. Thresholds are the points at which the probability of opting for one Likert category is equal to that of the prior adjacent category, or the 0.5 probability level. Also within the rating scale equation, ln represents the normal log, making it a logistic formula. P_{nik} represents the probability that a person *n* on an item *i* is observed in a rating category *k,* while $P_{ni(k-1)}$ represents the probability that this same person is observed in the category just prior to *k.* Thus, the rating scale formula represents the log of the odds of responding with respect to the adjacent categories of the scale (Wright and Mok, 2004). Per the rating scale equation, the rating scale model describes how the probability of a person responding to an item category is a logistic function of the relative distance on a linear scale between the respondent (person) measure location (B_n), the item difficulty measure location (D_i), and the 0.5 probability point threshold difficulty (F_k) for choosing between adjacent rating categories of the item.

This formula allows a linear transformation of the ordinal raw scores to derive person measures (B_n), item difficulties (D_i), and point threshold difficulties (F_k) in log-odds units referred to as logits. These logits are the units of a Rasch ruler. The common logit scale for item measures and person measures allows items and persons to be directly compared in a valid and meaningful way. Graphic depictions of Rasch rulers show the distributions of items and persons together, and are commonly referred to as variable maps, which are depicted in figures 8 through 13 and discussed below.

Calibrations and Diagnostics. Table 1 describes the eight categories of diagnostics examined within this study and the questions addressed with each diagnostic tool. Two of these diagnostic categories, item and person model fit, were examined using standardized (Zstd) and Means Square (Mnsq) infit and outfit indices. Zstd fit values within the -2.0 to 2.0 thresholds were considered as fitting the measurement model. Infit is a weighted index that is most sensitive to typical values while outfit is an unweighted index that is sensitive to extreme scores. Diagnostic categories including reliability, separation, sample targeting, person fit, item fit, dimensionality, item polarity, and category functioning were graphically represented in figures 2 through 33 and tables 2 and 3, including summary statistics output, item statistics output, item-person maps, item pathway plots, and category probability curves and output. These findings correspond with each of the six calibrations and the eight categories of measurement diagnostics used for this analysis (Table 1).

The six calibrations used are briefly described, then overall findings and interpretations across these calibrations are summarized. *Calibration 1:* The MSTAT-I, 22 item instrument was analyzed. *Calibration 2:* The MSTAT-II, 13 item instrument was analyzed to determine whether measurement characteristics improved from the prior version of the instrument. *Calibration 3*: Misfitting participants from calibration 2 were deleted for this additional calibration of the 13 item MSTAT-II including persons 49, 32, 74, 75, 2, 78, 99, and 103 based on the second calibration fit values. *Calibration 4*: A 9 item version was analyzed using all seven categories of the scale. Items eliminated from the 13 item scale included those with low polarity (< .50 point measure correlation) or high underfit (> 2.0 Zstd). Those eliminated included items 2, 5, 10, and 12 from the MSTAT-II. *Calibration 5:* The 9 item version was analyzed using only 5 categories that were collapsed from the 7 categories used after examining overlapping categories. *Calibration 6:* A five item version was analyzed using 5 categories that were collapsed from the 7 categories used. These five items were the *general ambiguity* items only (items 1, 3, 7, 11, and 13 from MSTAT-II) to provide a comparison of measurement data with only this stimulus type included.

Figures 2 through 7 and table 2 show reliability and separation data indicating levels supporting the differentiation of two groupings of person data across calibrations of instrument versions. Two or more groupings is favorable for differentiating people on the measure. Reliability/Separation levels declined, though very little, with reduction to 13 and 9 item instruments. The expected cost of reducing items was relatively minimal with respect to person reliability.

Tables 2 through 7 also provide summary statistics that include totaled raw scores (column 1) and Rasch measures (columns 3). Averaging the MSTAT-I raw score person total (calibration 1) provided a mean of 4.12 (SD = .71) which serves as a reference relative to the 7 levels of the Likert scale, though such average raw scores should not to be confused with true measures. However these raw score data can show that as a group, participants tended to be almost evenly divided on ambiguity tolerance. The MSTAT-I distribution was relatively normal. Cronbach alpha reliability for MSTAT-I = .86. On the Rasch person measures constructed from raw scores the normal distribution on the MSTAT-I had a mean measure of .07 (SD = .51) range of 3.43 logits. Though the item measure mean is calibrated to 0.0 (SD = .41), the items range was 1.99 logits on endorsement difficulty, making it a narrower distribution than that of items.

Sample targeting across all calibrations (figures 8-13) indicates a suitable instrument match to this group of educators. The difficulty level of the instrument's items corresponded with the participant measure locations, which helps minimize measurement error when compared with a poorly targeted sample. Sample targeting redundancy declined with 13 item MSTAT-IIversion but distribution coverage decline was minimal so overall the reduction in items with MSTAT-II did not sacrifice targeting very much. However, by reducing to and instrument with 9 items and then 5 items, a reduced range in distribution coverage resulted, as expected by the reduced diversity of items. Sample targeting overlap was diminished somewhat by the further reduction in items.

Person fit analysis of MSTAT-I (Figure 14) and MSTAT-II supported removal of eight extremely underfitting persons for calibration 3, which yielded improved item fit overall (Figures 15 and 16), but under-fit of one item (#5) persisted. The 9 item version resulted in improved item fit (Figure 17) compared with the 13 item MSTAT-II though two items were under-fitting slightly (#8 and #10). The five item calibration yielded the strongest item fit, as expected. In general, by eliminating items that did not support the dimension well, the measurement became more concentrated on the dimension and corresponding cohesion of the items. Collapsing to 5 categories the original 7 categories further reduced the level item underfit (Figures 18 and 19).

Principal components of residuals was used to examine dimensionality (Table 3). The analyses of 13 and 9 item instrument versions indicated a dominant dimension but also some evidence for secondary dimensions. With 39.0% and 48.3% of the variance accounted for by measures for the 13 item and 9 item instruments respectively, these shorter instrument versions allowed more cohesive AT measurement than the original 22 item MSTAT-I wherein 35.5% of the variance was accounted for by measures. First contrasts revealed relatively smaller unexplained variance though further examination of this unexplained variance is needed in subsequent studies.

Polarity of items improved in strength with the MSTAT-II and removal of misfitting persons (Figures 20-22). Further improvement was seen with reduced item versions that better represented the dominant dimension (figures 23-25), though the relative size of a secondary dimension increased with the five item instrument calibration.

Collapsing to 5 categories from the original 7 categories of the Likert scale was supported by category function analysis (Figures 26-33). Measure and threshold ordering was maintained, and overlap among categories was reduced, though not entirely, as illustrated on figures 31-33.

Limitations. Within this analysis the six calibrations were conducted using one modest sized sample of college level participants (n = 114) who all had career goals within teaching fields. Calibrations with data from other and larger samples of educator candidates and other groups could be used to corroborate findings and resulting interpretations across a more diverse population. Although fit analysis was used to help identify instrument mis-use such as careless responding, the elimination of all instrument misuse effects may not be possible because some instances of misuse may be undetected. Similarly, although McLain (2009) presented data supporting a low and non-significant relationship between MSTAT-II and a measure of social desirability, as with many self-rating approaches it remains possible that some form of socially desirable responding affected ratings and measurement of AT using these items despite the intent to create items that would not encourage socially desirable responding responding.

This investigation described only some of the crucial, fundamental Rasch diagnostic tools available for examining the measurement process. Further analyses of fit, dimensionality, and category functioning were not discussed here. Analyses such as differential item functioning (DIF), and factor sensitivity were also not covered within this paper.

Application of Findings. The 13 item MSTAT–II was supported by Rasch measurement findings toward valid assessment of AT among pre-service educators. Reliability and separation levels were appropriate, strong, and only modestly smaller than those of the longer 22 item version (Rasch person reliability was .84 for MSTAT-II versus .87 for MSTST-I). Targeting showed a favorable range of items to people and was only partially affected by the reduction in items due to the redundancy in locations of many of the items removed. That is, redundancy permitted some items to be expendable without changing the overlap of items and persons greatly. Polarity, item fit, and dimensionality were all improved with the MSTAT-II by comparison, particularly following elimination of under-fitting persons whose measures were extremely inconsistent and thereby difficult to interpret. Category functioning was characterized by measures and thresholds within the appropriate increasing direction for each level but categories overlapped excessively with seven levels used. These Rasch analytic findings build upon McLain's (2009) factor analysis and regression analysis investigation of business students and emergency medical technicians.

Examination of a shorter 9 item version of the AT instrument was also promising in that evidence for a single dimension was stronger than that of the 13 or 22 item instruments, while other measurement characteristics were also improved by the elimination of additional items. The instrument was more focused in its targeting of items to people, while retaining items across three of four stimulus types including *general ambiguity stimuli, complex stimuli*, and *insoluble stimuli*. Item 12 was the remaining item that was explicitly an *uncertain stimuli* item, though uncertainty can also be interpreted as overlapping the meaning within *general ambiguity* items without explicit use of the term 'uncertain' within items. Through a second calibration of the 9 item instrument, the reduction in rating categories from 7 to 5 or possibly fewer was supported for future administrations and empirical testing with a modified instrument. Related to this reduction in categories, using and even number such as 4 rating categories would eliminate the often overused middle category which perhaps requires respondents to use a lesser degree of thought in some cases. As Wolfe and Smith (2007) have noted in favor of an even number of categories "…the middle category is often used as a 'dumping ground' for participants that are compelled to provide a response but would not do so otherwise (pp. 231-232)."

Although the five item version yielded favorable fit and dimensionality because of its emphasis on general ambiguity, the range of items was considerably limited. While calibration of these five *general ambiguity* items was conducted primarily for a relative comparison to those of the other five calibrations, this five item version of general AT could be of use in studies where *complexity* and *insolubility* were not crucial, and the need to minimize items rated existed.

The range of pre-service teacher participant measures was relatively wide and normally distributed on AT. Considering this diversity found among the sample participants, as an intellectual disposition construct within studies of teacher candidate characteristics, AT has interesting potential to help provide meaningful insights on individual differences that have consequences for student learning. In light of prior AT research to date on teaching related tendencies regarding cognition, problem solving, and instructional approaches used, a measurement instrument such as the MSTAT-II and the shorter 9 item version appear to be valuable research tools toward better understanding differences among pre-service teachers that could shed light on means to addressing their needs within the teacher education process. That is, by carefully measuring whether an educator tends toward an aversion to ambiguity or to an attraction to ambiguity, for instance, meaningful investigations regarding the many possible implications of these educator tendencies for the thinking, problem solving, and learning among their pupils become more feasible.

References

Andrich D. (1978) A rating scale formulation for ordered response categories. *Psychometrika, 43,* 561-573.

Bohlig, M., Fisher, W.P. Jr., Masters, G.N., & Bond, T. (1998). Content validity and misfitting items. *Rasch Measurement Transactions,12*(1), 607.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences, third edition*. Mahwah NJ: Lawrence Erlbaum Associates.

Budner, J. (1962). Tolerance of ambiguity as a personality variable. Journal of Personality, 30, 29-40.

Cavanagh, R. F. (2012). Establishing validity of rating scale instrumentation in learning environment investigations. In R.F. Cavanaugh & R.F. Waugh (Eds), *Applications of Rasch measurement in learning environments research* (pp. 101-118). Rotterdam: Sense Publishers.

Curtis, D. D. (2004). Person misfit in attitude surveys: Influences, impacts, and implications. *International Education Journal, 5*(2), 125-143.

De Ayala, R.J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford.

Endres, M.L., Campt, R., & Milner, M. (2015). Is ambiguity tolerance malleable? Experimental evidence with potential implications for future research. *Frontiers in Psychology, 6*(619), 1-7.

Engelhard, G. (2013). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*. New York: Routlege.

Frenkel-Brunswik, E. (1948). Tolerance for ambiguity as a personality variable. *American Psychologist, 3*, 268. [Abstract Only]

Frenkel-Brunswik, E. (1949). Intolerance for ambiguity as an emotional and perceptual personality variable. *Journal of Personality, 18*, 108-143.

Gottlieb, A. S. (2006). Constructivist middle level teachers: An explorative analysis of their backgrounds, passions, and tolerance for ambiguity. (Doctoral Dissertation, Fordham University). *ProQuest Dissertation and Theses, UMI No. 3210267.*

Jonassen, D. H. & Grabowski, B. L. (1991). *Handbook of individual differences, learning, and instruction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lauriola, M., Foschi, R., Mosca, O., Weller, J. (2016). Attitude toward ambiguity: Empirically robust factors in self -report personality. *Assessment, 23*(3), 353-373.

Linacre, J.M. (2016a). *Winsteps® (Version 3.92.1) [Computer Software].* Beaverton, Oregon: Winsteps.com. Available from http://www.winsteps.com/

Linacre, J. M. (2016b). *Winsteps® Rasch measurement computer program User's Guide.* Beaverton, Oregon: Winsteps.com

McLain, D. L. (2009). Evidence of the properties of an ambiguity tolerance measure: The multiple stimulus types ambiguity tolerance scale-II (MSTAT-II). *Psychological Reports, 105*, 975-988.

McLain, D. L. (1993). The MSTAT-I: A new measure of an individual's tolerance for ambiguity. *Educational and Psychological Measurement, 53*, 183-189.

McLain, D.L., Kefallontis, E., & Armani, K. (2015). Ambiguity tolerance in organizations: definitional clarification and perspectives on future research. *Frontiers in Psychology, 6*(344), 1-7.

Messick, S. (1995). Validity of psychological assessment. *American Psychologist, 50*(9), 74-149.

Meyer, J. P. (2014). *jMetrik (Version 3.1) [Computer Software].* jMetrik.com. Retrieved April 26, 2013.

Moore, C. A. (1973). Styles of teacher behavior under simulated teaching conditions (Doctoral Dissertation, Stanford University). *Dissertation Abstracts International, 34*, 3149A-3150A. (University Microfilms No. 73-30,449).

Peters, W. H. & Amburgey, B. S. (1982). Teacher intellectual disposition and cognitive classroom verbal interactions. *Journal of Educational Research, 76*, 94-99.

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

Ream, M. N. (1984). Effects of teaching strategies in increasing tolerance of ambiguity among middle school students (Doctoral Dissertation, West Virginia University). *Dissertation Abstracts International, 45*(5A), 1289.

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

Reise, S. P., Ainsworth, A. T., & Haviland, M. G. (2005). Item response theory: Fundamentals, applications, and promise in psychological research. *Current Directions in Psychological Science, 14*(2), 95-101.

Rittschof, K. A. & Chambers, W. L. (2011). Modern measurement information graphics for understanding student performance differences. *International Journal of Teaching and Learning in Higher Education, 23*(3), 350-363.

Schommer-Aikins, M. & Hutter, R. (2002). Epistemological beliefs and thinking about everyday controversial issues. *Journal of Psychology, 136*(1), 5-20.

Schuman, H. & Presser, S. (1996). *Questions and answers in attitude surveys: Experiments on question form, wording, and context.* Thousand Oaks, CA: Sage Publications.

Smith, E. V. & Smith, R. M. (2004). *Introduction to Rasch measurement: Theory, models, and applications*. Maple Grove, MN: JAM Press.

Williams, R. A. (1980). The relationship between field independence, tolerance of ambiguity and stress in undergraduate nursing students (Doctoral Dissertation, University of Washington). *Dissertation Abstracts International, 41(07)*, 3121.

Wolfe, E. W., and E. V. Smith, Jr. 2007. Instrument development tools and activities for measure validation using Rasch Models: Part I – Instrument development tools. In *Rasch measurement: Advanced and specialized applications,* eds. E. Smith, Jr., and R. M. Smith, 202-242. Maple Grove, MN: Jam Press.

Wright, B. D., and M. C. Mok. (2004). An overview of the family of Rasch measurement models. In *Introduction to Rasch measurement: Theory, models, and applications*, eds. E. V. Smith, and R. M. Smith, 1-24. Maple Grove, MN: JAM Press.

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

Table 1

Rasch Model Diagnostic Tools for Improving Rating Scale Measurement

Diagnostic Tool	General and Specific Questions and Criteria Regarding Unidimensional Measurement
Reliability	*General:* Do person measures indicate a broad enough ability range and a sufficient number of items to be reproducible and do item measures represent a broad enough difficulty range and a sufficient number of students to be reproducible? *Specific:* Is the person reliability level .8 or above and is the item reliability .9 or above?
Separation	*General:* Do person measures allow discrimination between at least two different levels of the scale (high and low ratings) and do item measures allow discrimination of at least three different levels of difficulty (high, medium, low)? *Specific:* Are separation index levels greater than 2.0 for persons and greater than 3.0 for items?
Sample Targeting	*General:* Do item measures and person measures correspond to one another? *Specific:* Are the distributions of item and person measures located within a shared range on the logit scale, and thereby matched on the variable map?
Person Fit	*General:* Do person measures function together as an overall measure of the modeled dimension? *Specific:* Are fit values for each person within the expected range (less than Zstd = 2.0) to avoid underfitting the model?
Item Fit	*General:* Do item measures function together as an overall measure of the modeled dimension? *Specific:* Are fit values for each item within the expected range (less than Zstd = 2.0) to avoid underfitting the model?
Dimensionality	*General:* Do the items of the instrument as a whole measure a primary, dominant dimension? *Specific:* Are the variances (percentages) accounted for relatively small for any non-primary dimensions and largest for the primary dimension (ideally above 50%)?
Item Polarity	*General:* Do items function in unison? *Specific:* Are point measure correlations positive and strong (ideally above .50)?
Category Functioning	*General:* Do constructed measures function in the expected manner relative to "more" and "less"? *Specific:* Are average measures and thresholds for each subsequent Likert category in advancing order, from smaller to larger?

1. I don't tolerate ambiguous situations very well. *(1. General Ambiguity)*

2. I find it difficult to respond when faced with an unexpected event. *(MSTAT-I Only)*

3. I don't think new situations are any more threatening than familiar situations. *(MSTAT-I Only)*

4. I'm drawn to situations that can be interpreted in more than one way. *(MSTAT-I Only)*

5. I would rather avoid solving a problem that must be viewed from several different perspectives. *(2. Insoluble)*

6. I try to avoid situations that are ambiguous. *(3. General Ambiguity)*

7. I am good at managing unpredictable situations. *(MSTAT-I Only)*

8. I prefer familiar situations to new ones. *(4 Unfamiliar)*

9. Problems that cannot be considered from just one point of view are a little threatening. *(5. Insoluble)*

10. I avoid situations that are too complicated for me to easily understand. *(6. Complex)*

11. I am tolerant of ambiguous situations. *(7. General Ambiguity)*

12. I enjoy tackling problems that are complex enough to be ambiguous. *(8. Complex)*

13. I try to avoid problems that don't seem to have only one "best" solution. *(9. Insoluble)*

14. I often find myself looking for something new, rather than trying to hold things constant in my life. *(MSTAT-I Only)*

15. I generally prefer novelty over familiarity. *(10 Unfamiliar)*

16. I dislike ambiguous situations. *(11. General Ambiguity)*

17. Some problems are so complex that just trying to understand them is fun. *(MSTAT-I Only)*

18. I have little trouble coping with unexpected events. *(MSTAT-I Only)*

19. I pursue problem situations that are so complex some people call them "mind boggling". *(MSTAT-I Only)*

20. I find it hard to make a choice when the outcome is uncertain. *(12. Uncertain)*

21. I enjoy an occasional surprise. *(MSTAT-I Only)*

22. I prefer a situation in which there is some ambiguity. *(13. General Ambiguity)*

Figure 1. MSTAT-I (McLain, 1993) items are numbered 1 through 22. MSTAT-II (McLain, 2009) item numbers and stimulus type are shown to the right of the corresponding items included in this more recent version.

Rasch Summary Statistics.

```
     SUMMARY OF 114 MEASURED Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      90.7      22.0        .07     .18      1.03    -.4   1.05    -.4 |
| P.SD      15.5        .2        .51     .01       .82    2.4    .86    2.4 |
| S.SD      15.6        .2        .51     .01       .82    2.4    .86    2.4 |
| MAX.     139.0      22.0       2.02     .27      5.44    6.6   6.02    6.9 |
| MIN.      47.0      21.0      -1.41     .17       .14   -5.1    .15   -5.0 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .21 TRUE SD     .46  SEPARATION  2.22  Person RELIABILITY  .83 |
|MODEL RMSE    .18 TRUE SD     .47  SEPARATION  2.65  Person RELIABILITY  .87 |
| S.E. OF Person MEAN = .05                                                   |
-------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .86  SEM = 5.79

     SUMMARY OF 22 MEASURED Item
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN     470.0     113.8        .00     .08      1.01    -.1   1.05     .1 |
| P.SD      61.7        .4        .40     .00       .31    2.4    .37    2.7 |
| S.SD      63.1        .4        .41     .00       .31    2.4    .38    2.8 |
| MAX.     666.0     114.0        .63     .10      1.77    5.1   2.11    6.9 |
| MIN.     367.0     113.0      -1.36     .08       .55   -4.4    .54   -4.5 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .08 TRUE SD     .39  SEPARATION  4.68  Item   RELIABILITY  .96 |
|MODEL RMSE    .08 TRUE SD     .39  SEPARATION  4.96  Item   RELIABILITY  .96 |
| S.E. OF Item MEAN = .09                                                     |
-------------------------------------------------------------------------------
Item RAW SCORE-TO-MEASURE CORRELATION = -1.00
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000
```

Figure 2. Calibration 1. MSTAT-I.

```
     SUMMARY OF 114 MEASURED Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE    S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      51.8      13.0        -.03     .24      1.02    -.3   1.03    -.3 |
| P.SD      10.0        .1         .60     .01       .90    1.9    .96    2.0 |
| S.SD      10.0        .1         .60     .01       .91    2.0    .97    2.0 |
| MAX.      79.0      13.0        1.77     .31      6.54    6.6   7.67    7.3 |
| MIN.      27.0      12.0       -1.57     .24       .06   -4.7    .06   -4.7 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .28 TRUE SD     .52  SEPARATION  1.86  Person RELIABILITY  .78 |
|MODEL RMSE    .24 TRUE SD     .54  SEPARATION  2.22  Person RELIABILITY  .83 |
| S.E. OF Person MEAN = .06                                                   |
-------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .82  SEM = 4.27

     SUMMARY OF 13 MEASURED Item
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE    S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN     454.4     113.8         .00     .08      1.00    -.3   1.03    -.1 |
| P.SD      46.5        .4         .31     .00       .34    2.5    .41    2.8 |
| S.SD      48.4        .4         .32     .00       .36    2.6    .42    2.9 |
| MAX.     528.0     114.0         .59     .08      1.96    6.0   2.21    7.2 |
| MIN.     367.0     113.0        -.49     .08       .66   -3.0    .66   -3.0 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .09 TRUE SD     .30  SEPARATION  3.43  Item   RELIABILITY  .92 |
|MODEL RMSE    .08 TRUE SD     .30  SEPARATION  3.65  Item   RELIABILITY  .93 |
| S.E. OF Item MEAN = .09                                                     |
-------------------------------------------------------------------------------
                DELETED:       9 Item
Item RAW SCORE-TO-MEASURE CORRELATION = -1.00
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000
```

Figure 3. Calibration 2. MSTAT-II.

```
     SUMMARY OF 106 MEASURED Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   S.E.       MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      51.7      13.0        -.08     .27      1.00    -.2   1.00    -.2 |
| P.SD       9.4        .1         .67     .01       .55    1.6    .55    1.6 |
| S.SD       9.4        .1         .67     .01       .56    1.6    .55    1.6 |
| MAX.      75.0      13.0        1.73     .32      2.20    2.6   2.20    2.6 |
| MIN.      31.0      12.0       -1.56     .26       .08   -4.4    .08   -4.4 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .30 TRUE SD     .60  SEPARATION  2.02  Person RELIABILITY  .80 |
|MODEL RMSE    .27 TRUE SD     .62  SEPARATION  2.30  Person RELIABILITY  .84 |
| S.E. OF Person MEAN = .07                                                   |
-------------------------------------------------------------------------------
                DELETED:       8 Person
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .83  SEM = 3.90

     SUMMARY OF 13 MEASURED Item
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   S.E.       MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN     421.2     105.8         .00     .09      1.00    -.2   1.00    -.2 |
| P.SD      43.2        .4         .37     .00       .33    2.3    .33    2.4 |
| S.SD      45.0        .4         .39     .00       .34    2.4    .35    2.5 |
| MAX.     489.0     106.0         .70     .10      1.85    5.2   1.86    5.2 |
| MIN.     341.0     105.0        -.59     .09       .64   -3.1    .64   -3.1 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .10 TRUE SD     .36  SEPARATION  3.62  Item   RELIABILITY  .93 |
|MODEL RMSE    .09 TRUE SD     .36  SEPARATION  3.86  Item   RELIABILITY  .94 |
| S.E. OF Item MEAN = .11                                                     |
-------------------------------------------------------------------------------
                DELETED:       9 Item
Item RAW SCORE-TO-MEASURE CORRELATION = -1.00
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000
```

Figure 4. Calibration 3. MSTAT-II, underfitting persons removed.

```
     SUMMARY OF 106 MEASURED Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT          OUTFIT    |
|          SCORE     COUNT     MEASURE   S.E.       MNSQ   ZSTD   MNSQ   ZSTD  |
|-------------------------------------------------------------------------------|
| MEAN      36.0       9.0        -.18     .36        .99    -.2    .99    -.2  |
| P.SD       7.2        .0         .95     .02        .64    1.5    .64    1.5  |
| S.SD       7.2        .0         .96     .03        .64    1.5    .64    1.5  |
| MAX.      53.0       9.0        2.42     .46       2.87    2.9   2.92    2.9  |
| MIN.      18.0       9.0       -2.45     .34        .07   -3.6    .07   -3.6  |
|-------------------------------------------------------------------------------|
| REAL RMSE    .41 TRUE SD     .86  SEPARATION  2.12  Person RELIABILITY  .82  |
|MODEL RMSE    .36 TRUE SD     .88  SEPARATION  2.42  Person RELIABILITY  .85  |
| S.E. OF Person MEAN = .09                                                     |
-------------------------------------------------------------------------------
                DELETED:      8 Person
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .83  SEM = 2.93

     SUMMARY OF 9 MEASURED Item
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT          OUTFIT    |
|          SCORE     COUNT     MEASURE   S.E.       MNSQ   ZSTD   MNSQ   ZSTD  |
|-------------------------------------------------------------------------------|
| MEAN     423.8     106.0         .00     .11        .99    -.1    .99    -.2  |
| P.SD      40.7        .0         .44     .00        .25    1.8    .24    1.8  |
| S.SD      43.2        .0         .47     .00        .26    1.9    .26    1.9  |
| MAX.     489.0     106.0         .89     .11       1.42    2.7   1.40    2.7  |
| MIN.     341.0     106.0        -.74     .10        .66   -2.9    .67   -2.8  |
|-------------------------------------------------------------------------------|
| REAL RMSE    .11 TRUE SD     .43  SEPARATION  3.89  Item   RELIABILITY  .94  |
|MODEL RMSE    .11 TRUE SD     .43  SEPARATION  4.10  Item   RELIABILITY  .94  |
| S.E. OF Item MEAN = .16                                                       |
-------------------------------------------------------------------------------
                DELETED:     13 Item
Item RAW SCORE-TO-MEASURE CORRELATION = -1.00
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000
```

Figure 5. Calibration 4. A 9-Item subset of MSTAT-II.

```
     SUMMARY OF 106 MEASURED Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   S.E.       MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      27.1       9.0       -.01     .40       1.01    -.2   1.02    -.1 |
| P.SD       6.8        .0       1.07     .07        .64    1.5    .66    1.5 |
| S.SD       6.8        .0       1.07     .07        .65    1.5    .66    1.5 |
| MAX.      43.0       9.0       3.21     .75       3.41    3.1   3.30    2.8 |
| MIN.      12.0       9.0      -2.58     .36        .08   -3.8    .08   -3.7 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .46 TRUE SD     .96  SEPARATION  2.10  Person RELIABILITY  .82 |
|MODEL RMSE    .40 TRUE SD     .99  SEPARATION  2.44  Person RELIABILITY  .86 |
| S.E. OF Person MEAN = .10                                                   |
-------------------------------------------------------------------------------
               DELETED:      8 Person
Person RAW SCORE-TO-MEASURE CORRELATION = .99
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .84  SEM = 2.73

     SUMMARY OF 9 MEASURED Item
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   S.E.       MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN     319.3     106.0        .00     .11       1.00    -.1   1.02     .1 |
| P.SD      36.4        .0        .46     .00        .22    1.6    .22    1.6 |
| S.SD      38.6        .0        .49     .00        .23    1.7    .23    1.7 |
| MAX.     380.0     106.0        .89     .12       1.40    2.7   1.39    2.5 |
| MIN.     249.0     106.0       -.78     .11        .68   -2.8    .67   -2.8 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .12 TRUE SD     .45  SEPARATION  3.77  Item   RELIABILITY  .93 |
|MODEL RMSE    .11 TRUE SD     .45  SEPARATION  3.96  Item   RELIABILITY  .94 |
| S.E. OF Item MEAN = .16                                                     |
-------------------------------------------------------------------------------
               DELETED:     13 Item
Item RAW SCORE-TO-MEASURE CORRELATION = -1.00
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000
```

Figure 6. Calibration 5. A 9-item subset of MSTAT-II with 5 categories collapsed from the 7 categories used.

```
     SUMMARY OF 104 MEASURED (NON-EXTREME) Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   S.E.       MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      14.6       5.0        -.17     .59        .99    -.3  1.00    -.3 |
| P.SD       3.9        .0        1.33     .09        .87    1.6   .88    1.6 |
| S.SD       3.9        .0        1.34     .09        .88    1.6   .88    1.6 |
| MAX.      24.0       5.0        3.76    1.06       3.38    2.7  3.37    2.7 |
| MIN.       6.0       5.0       -3.47     .54        .07   -3.2   .07   -3.1 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .69 TRUE SD    1.14  SEPARATION  1.65  Person RELIABILITY  .73 |
|MODEL RMSE    .60 TRUE SD    1.19  SEPARATION  1.99  Person RELIABILITY  .80 |
| S.E. OF Person MEAN = .13                                                   |
-------------------------------------------------------------------------------
  MAXIMUM EXTREME SCORE:      2 Person 1.9%
               DELETED:      8 Person

     SUMMARY OF 106 MEASURED (EXTREME AND NON-EXTREME) Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   S.E.       MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      14.8       5.0        -.07     .62                                |
| P.SD       4.1        .0        1.50     .19                                |
| S.SD       4.1        .0        1.50     .19                                |
| MAX.      25.0       5.0        5.05    1.85                                |
| MIN.       6.0       5.0       -3.47     .54                                |
|-----------------------------------------------------------------------------|
| REAL RMSE    .73 TRUE SD    1.31  SEPARATION  1.80  Person RELIABILITY  .76 |
|MODEL RMSE    .65 TRUE SD    1.35  SEPARATION  2.09  Person RELIABILITY  .81 |
| S.E. OF Person MEAN = .15                                                   |
-------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = .98
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .78  SEM = 1.91

     SUMMARY OF 5 MEASURED (NON-EXTREME) Item
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   S.E.       MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN     314.6     106.0         .00     .13        .99    -.1  1.00     .0 |
| P.SD      21.6        .0         .35     .00        .15    1.1   .17    1.3 |
| S.SD      24.1        .0         .39     .00        .16    1.3   .19    1.4 |
| MAX.     340.0     106.0         .54     .13       1.18    1.3  1.21    1.5 |
| MIN.     281.0     106.0        -.41     .13        .74   -2.1   .72   -2.2 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .13 TRUE SD     .32  SEPARATION  2.47  Item   RELIABILITY  .86 |
|MODEL RMSE    .13 TRUE SD     .32  SEPARATION  2.54  Item   RELIABILITY  .87 |
| S.E. OF Item MEAN = .17                                                     |
-------------------------------------------------------------------------------
               DELETED:     17 Item
Item RAW SCORE-TO-MEASURE CORRELATION = -1.00
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000
```

Figure 7. Calibration 6. A 5-item subset of MSTAT-II using only *general ambiguity* stimuli items and 5 categories.

Table 2

Rasch Person (Test) Reliability and Separation Indices

Calibration/Items	Reliability	Separation
1) 22	.87	2.65
2) 13	.83	2.22
3) 13	.84	2.30
4) 9	.85	2.42
5) 9	.86	2.44
6) 5	.81	2.09

Summary of person reliability and separation across six calibrations.

Rasch Sample Targeting Variable Maps.

MSTAT-I (22 Item) **MSTAT-II (13 Item)**

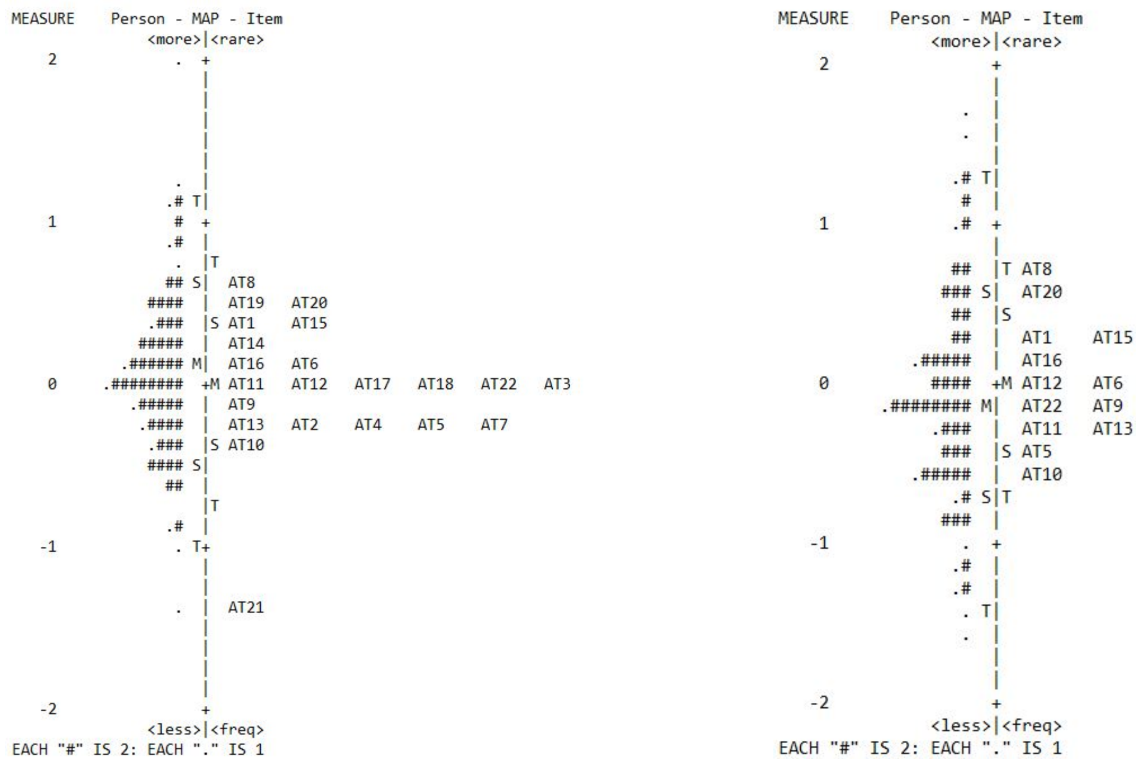

Figure 8. Item-Person Variable Maps using MSTAT-I and MSTAT-II compared.

9 Item Instrument

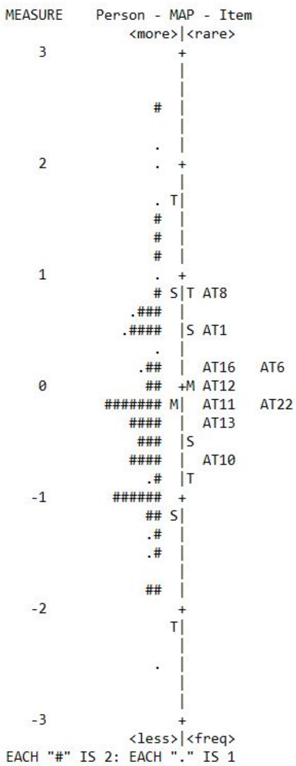

Figure 9. Item-Person Variable Map using 9 item version derived from MSTAT-II.

5 Item Instrument

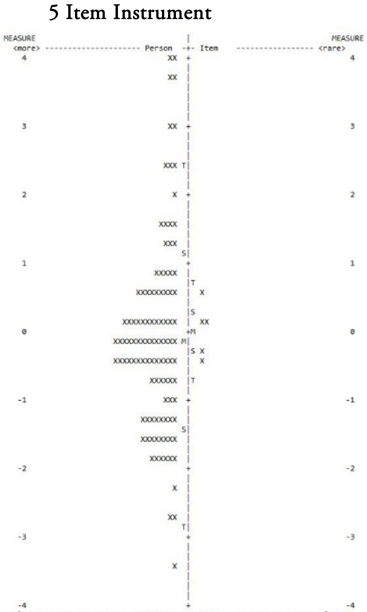

Figure 10. Item-Person Variable Map using 5 item version derived from MSTAT-II.

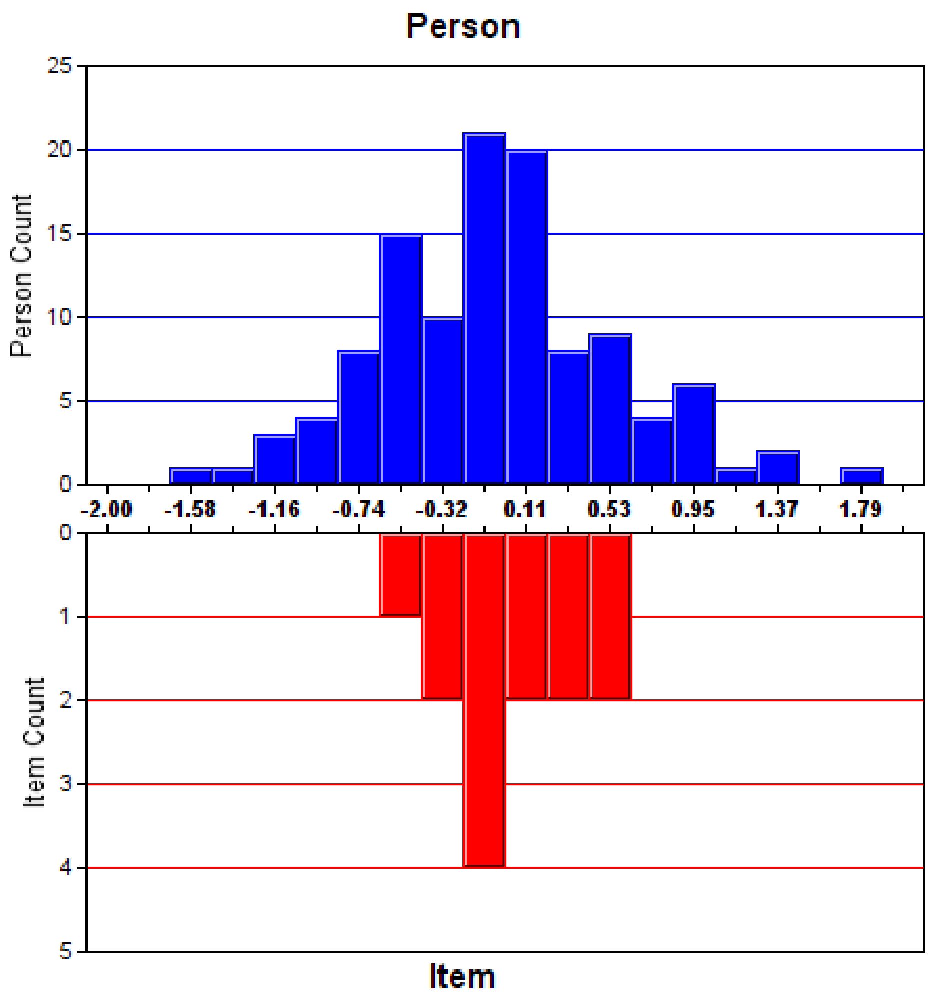

Figure 11. Histogram Variable Map using MSTAT-II, 13 Item Instrument for sample targeting graphic visualization.

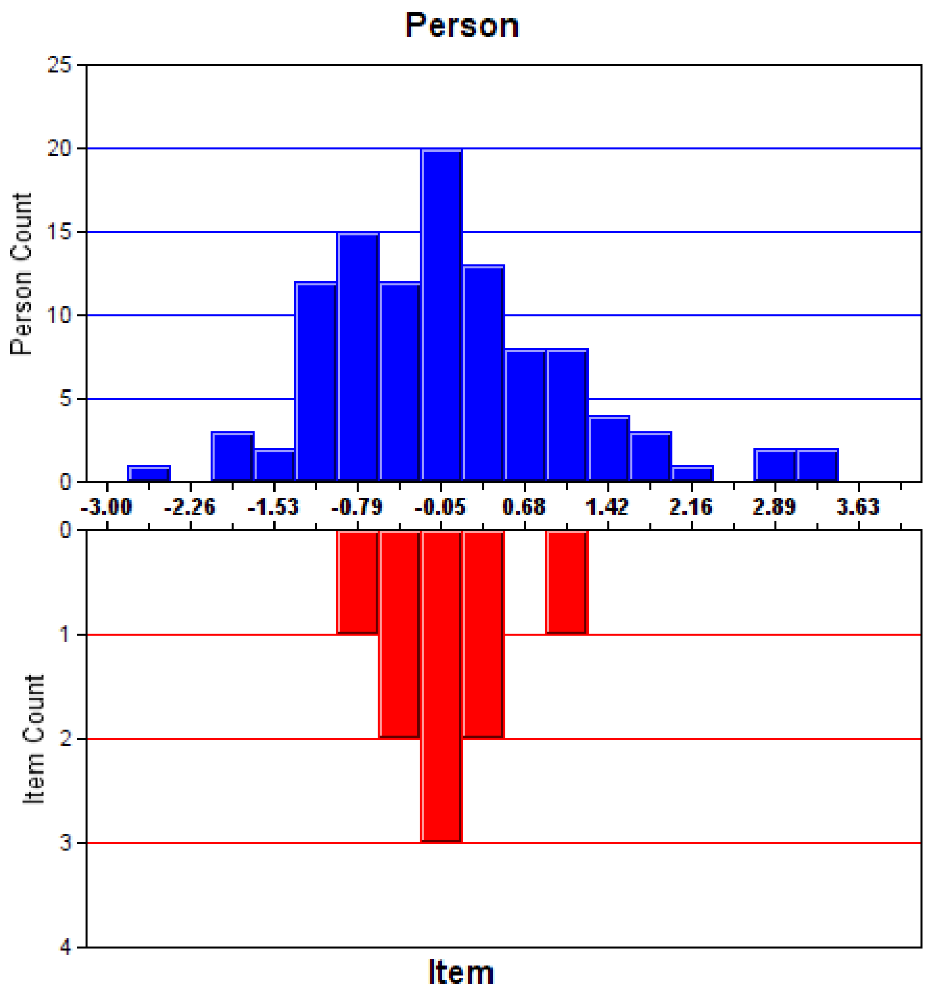

Figure 12. Histogram variable map using 9 item instrument subset of MSTAT-II for sample targeting graphic visualization.

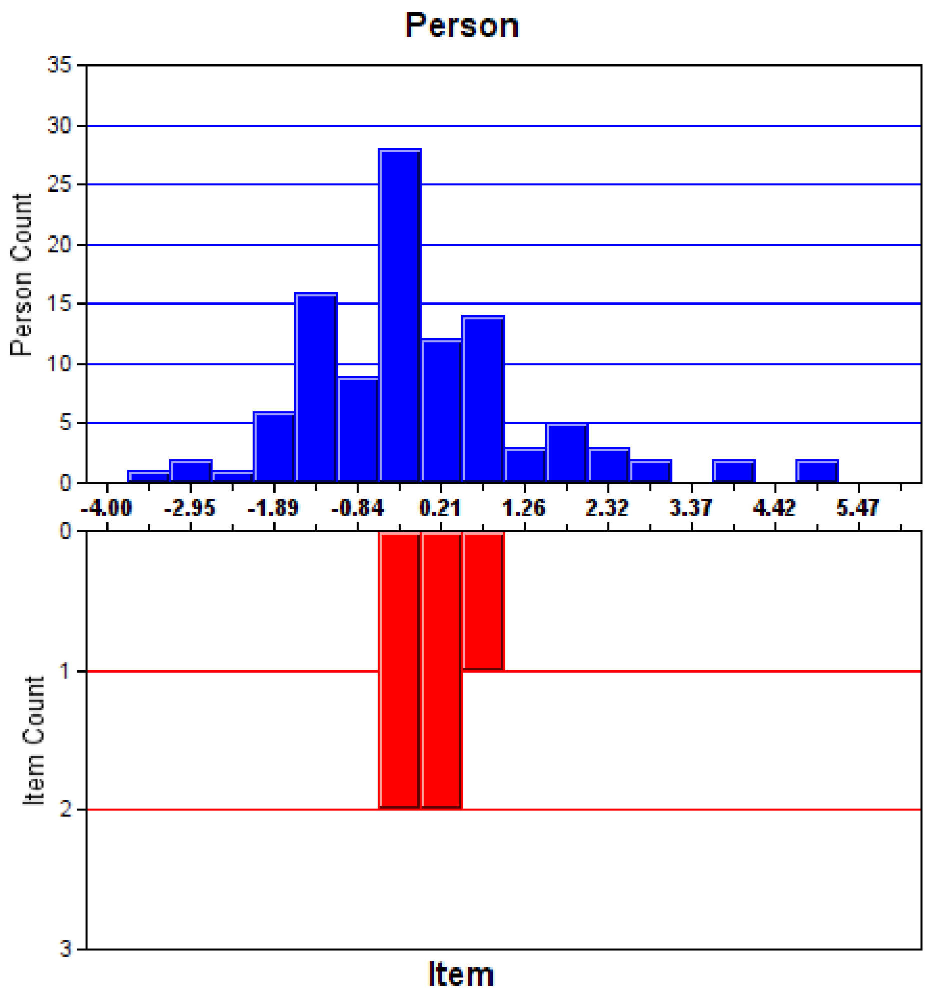

Figure 13. Histogram variable map using 5 Item instrument subset (general ambiguity stimuli) from MSTAT-II for sample targeting graphic visualization.

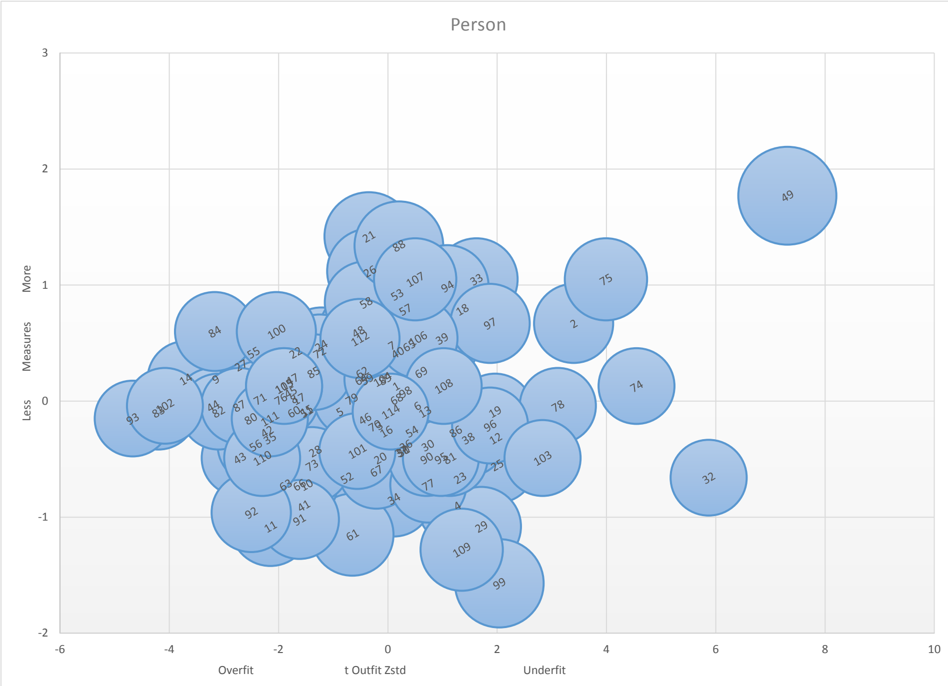

Figure 14. Person Outfit using MSTAT-I

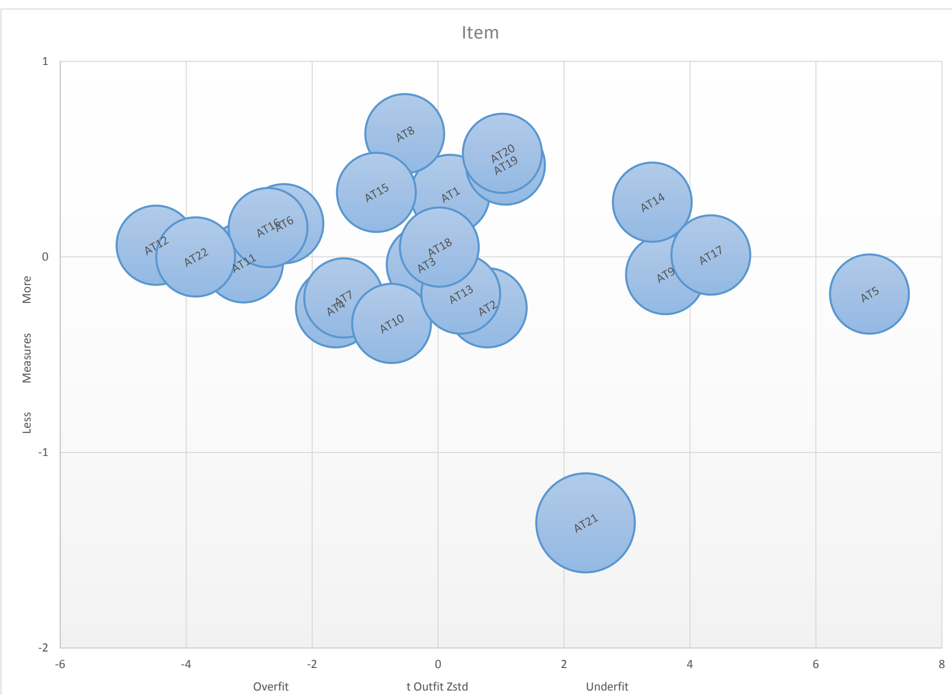

Figure 15. Item Outfit using MSTAT-I

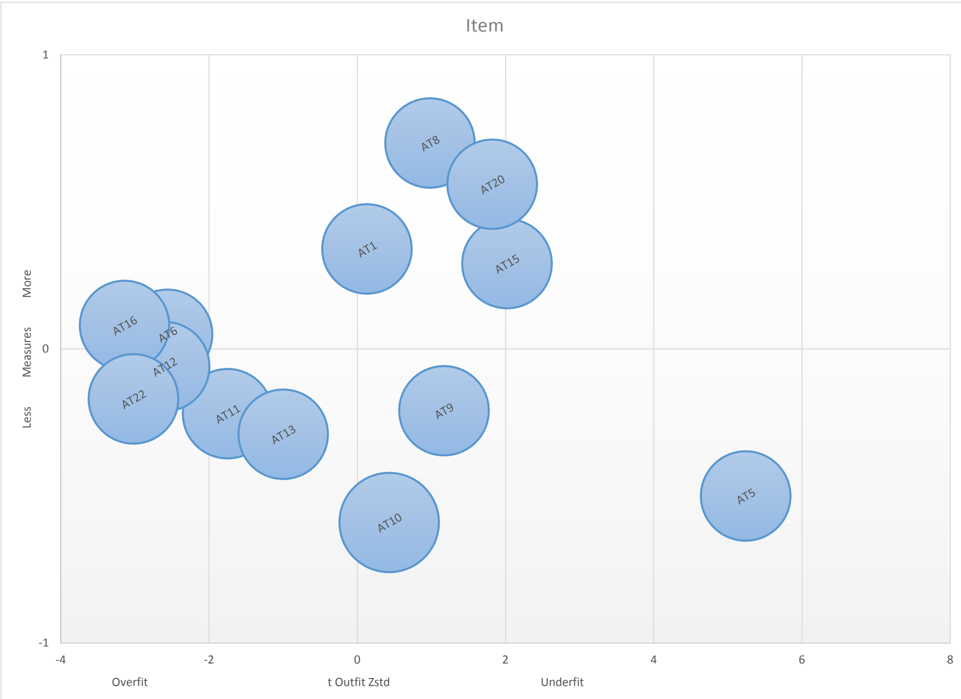

Figure 16. Item Outfit using MSTAT-II

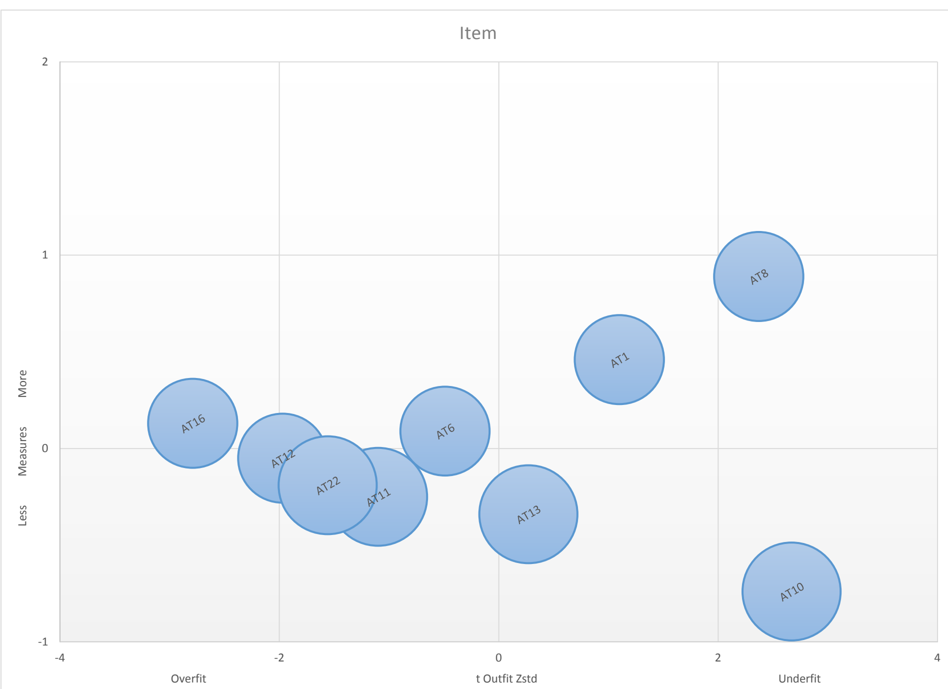

Figure 17. Item Outfit using 9 item, 7 category version derived from MSTAT-II

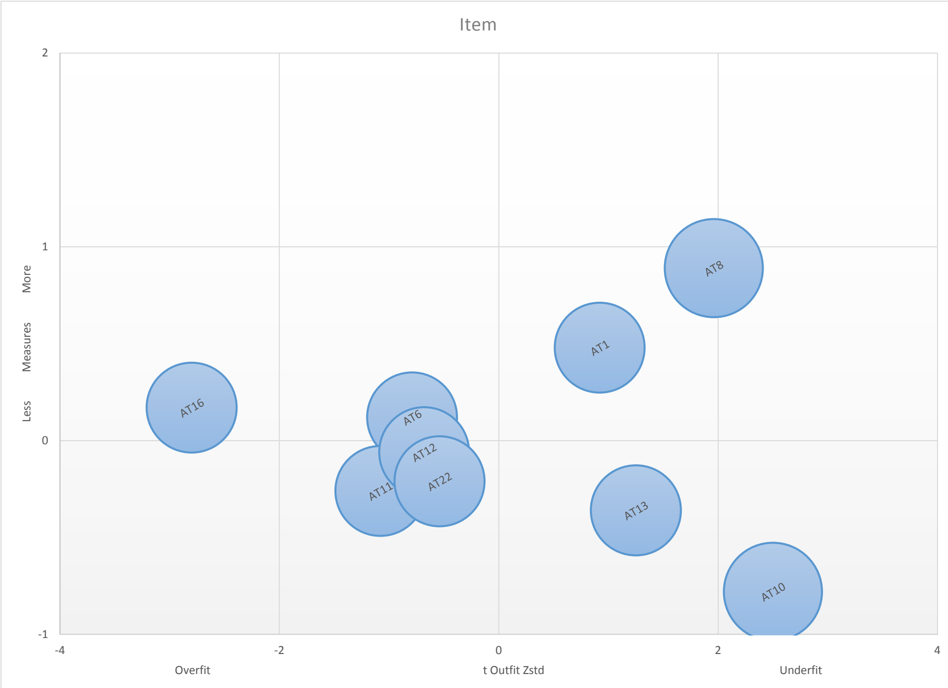

Figure 18. Item Outfit using 9 item, 5 category version derived from MSTAT-II

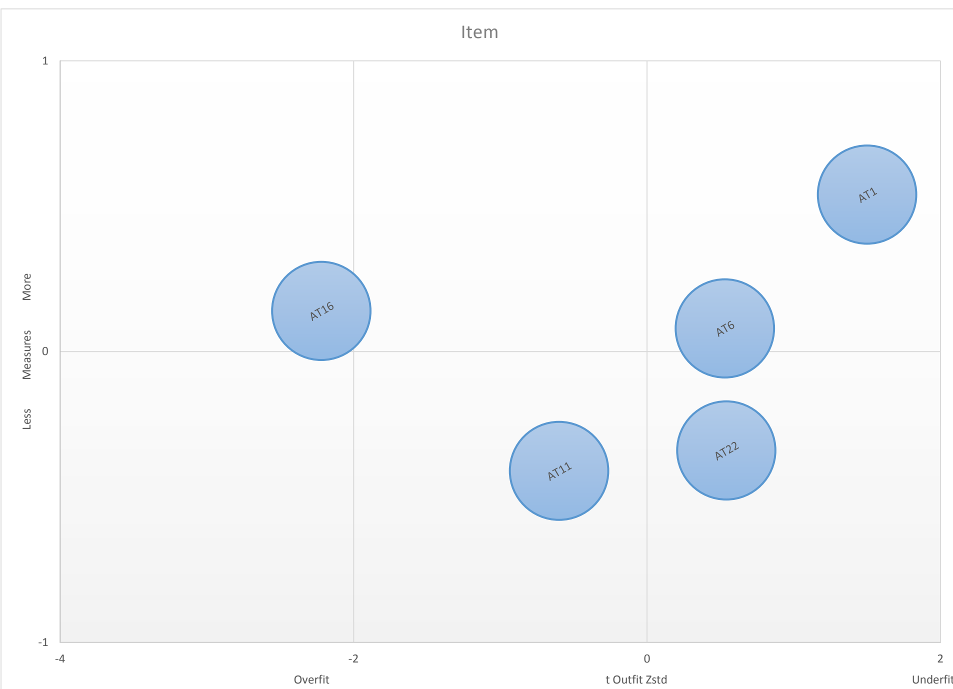

Figure 19. Item Outfit using 5 item, 5 category version derived from MSTAT-II

Table 3

Dimensionality Analysis with Principal Components of Residuals across Calibrations.

Calibration / Items	Measures	First Contrast
1) 22	35.5%	9.3%
2) 13	37.1%	9.9%
3) 13	39.0%	9.6%
4) 9	48.1%	9.6%
5) 9	48.3%	8.9%
6) 5	52.6%	15.6%

Note. Percentage of variance accounted for by measures and by the first contrast within principal components analysis (PCA) of residuals. Higher percentages for measures provide stronger support for unidimensionality (e.g. calibrations 4 through 6). First contrasts of residuals indicate strength of an additional dimension. Percentages of variance calculated from PCA eigenvalues.

Polarity Analysis of Items and corresponding Item Statistics

Item STATISTICS: CORRELATION ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
5	505	114	-.19	.08	1.77	5.1	2.11	6.9	.31	.50	24.6	31.4	AMBIG5
21	666	114	-1.36	.10	1.20	1.4	1.36	2.3	.32	.41	37.7	38.0	AMBIG21
9	483	113	-.09	.08	1.21	1.6	1.51	3.6	.33	.51	36.3	31.0	AMBIG9
14	426	114	.28	.08	1.45	3.3	1.47	3.4	.34	.52	30.7	30.6	AMBIG14
18	465	114	.05	.08	.99	.0	1.00	.0	.41	.51	31.6	30.7	AMBIG18
17	471	114	.01	.08	1.61	4.3	1.63	4.3	.44	.51	21.9	30.8	AMBIG17
13	505	114	-.19	.08	1.00	.1	1.04	.4	.46	.50	37.7	31.4	AMBIG13
15	413	113	.33	.08	.88	-1.0	.88	-1.0	.46	.52	37.2	30.5	AMBIG15
20	383	114	.53	.08	1.12	1.0	1.12	1.0	.50	.52	38.6	30.7	AMBIG20
7	508	114	-.21	.08	.83	-1.4	.82	-1.5	.51	.50	42.1	31.4	AMBIG7
3	475	113	-.04	.08	.97	-.2	.97	-.2	.53	.51	31.9	30.9	AMBIG3
2	516	114	-.26	.08	1.11	.9	1.09	.8	.55	.50	32.5	31.6	AMBIG2
10	528	114	-.34	.08	.93	-.5	.90	-.7	.55	.50	34.2	32.1	AMBIG10
22	473	114	.00	.08	.59	-3.9	.59	-3.8	.55	.51	42.1	30.8	AMBIG22
19	391	113	.47	.08	1.14	1.1	1.13	1.1	.56	.52	29.2	30.6	AMBIG19
1	419	114	.32	.08	1.02	.2	1.02	.2	.57	.52	31.6	30.5	AMBIG1
11	477	114	-.03	.08	.65	-3.2	.66	-3.1	.59	.51	45.6	30.8	AMBIG11
6	444	114	.17	.08	.73	-2.4	.73	-2.4	.63	.52	41.2	30.6	AMBIG6
4	516	114	-.26	.08	.82	-1.5	.81	-1.6	.63	.50	43.0	31.6	AMBIG4
8	367	114	.63	.08	.92	-.6	.93	-.5	.64	.51	35.1	31.0	AMBIG8
12	463	114	.06	.08	.55	-4.4	.54	-4.5	.68	.51	52.6	30.7	AMBIG12
16	447	114	.15	.08	.70	-2.7	.70	-2.7	.68	.51	43.9	30.6	AMBIG16
MEAN	470.0	113.8	.00	.08	1.01	-.1	1.05	.1			36.4	31.3	
P.SD	61.7	.4	.40	.00	.31	2.4	.37	2.7			7.0	1.5	

Figure 20. Calibration 1, MSTAT-I instrument. Point measure correlation (column 10, 'PTMEASUR') indicates item polarity. Item statistics are sequenced in order of correlation. Item fit statistics are shown in columns 6 through 9 as standardsized (Zstd) and means square (Mnsq) indices of weighted infit and unweighted outfit.

Item STATISTICS: CORRELATION ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
5	505	114	-.33	.08	1.96	6.0	2.21	7.2	.34	.55	24.6	33.7	AT5
15	413	113	.25	.08	1.10	.8	1.13	1.0	.39	.56	40.7	33.1	AT15
9	483	113	-.21	.08	1.26	1.9	1.43	3.0	.42	.56	39.8	33.6	AT9
22	473	114	-.12	.08	.70	-2.6	.71	-2.5	.53	.56	40.4	33.5	AT22
13	505	114	-.33	.08	1.02	.2	1.03	.3	.53	.55	37.7	33.7	AT13
20	383	114	.48	.08	1.16	1.3	1.15	1.2	.56	.56	37.7	32.8	AT20
10	528	114	-.49	.08	.94	-.5	.93	-.5	.61	.55	32.5	33.8	AT10
1	419	114	.24	.08	1.04	.4	1.04	.4	.62	.56	33.3	33.1	AT1
12	463	114	-.05	.08	.67	-2.9	.68	-2.8	.64	.56	49.1	33.4	AT12
8	367	114	.59	.08	1.01	.1	1.00	.0	.64	.56	36.0	32.7	AT8
11	477	114	-.15	.08	.66	-3.0	.66	-3.0	.65	.56	42.1	33.5	AT11
6	444	114	.07	.08	.72	-2.4	.72	-2.4	.69	.56	50.0	33.3	AT6
16	447	114	.06	.08	.69	-2.7	.69	-2.8	.74	.56	51.8	33.3	AT16
MEAN	454.4	113.8	.00	.08	1.00	-.3	1.03	-.1			39.7	33.3	
P.SD	46.5	.4	.31	.00	.34	2.5	.41	2.8			7.3	.3	

Figure 21. Calibration 2, MSTAT-II instrument. Point measure correlation (column 10, ‘PTMEASUR’) indicates item polarity. Item statistics are sequenced in order of correlation. Item fit statistics are shown in columns 6 through 9 as standardsized (Zstd) and means square (Mnsq) indices of weighted infit and unweighted outfit.

Item STATISTICS: CORRELATION ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
15	383	105	.29	.09	1.24	1.7	1.28	2.0	.36	.57	41.9	35.9	AT15
5	479	106	-.50	.09	1.85	5.2	1.86	5.2	.47	.56	30.2	37.0	AT5
20	357	106	.56	.09	1.25	1.8	1.25	1.8	.50	.58	38.7	36.0	AT20
22	442	106	-.17	.09	.65	-3.0	.65	-3.0	.52	.57	44.3	36.4	AT22
9	442	105	-.21	.09	1.14	1.0	1.16	1.2	.53	.57	41.9	36.5	AT9
10	489	106	-.59	.10	1.06	.5	1.05	.4	.58	.56	34.9	37.2	AT10
11	447	106	-.22	.09	.79	-1.7	.78	-1.7	.60	.57	46.2	36.5	AT11
13	455	106	-.29	.09	.87	-1.0	.87	-1.0	.61	.57	39.6	36.7	AT13
8	341	106	.70	.09	1.14	1.0	1.13	1.0	.62	.58	38.7	36.0	AT8
1	382	106	.34	.09	1.01	.1	1.01	.1	.62	.58	32.1	35.9	AT1
12	429	106	-.06	.09	.69	-2.6	.69	-2.6	.63	.57	52.8	36.2	AT12
6	416	106	.05	.09	.69	-2.6	.70	-2.6	.70	.57	52.8	36.2	AT6
16	413	106	.08	.09	.64	-3.1	.64	-3.1	.77	.57	54.7	36.2	AT16
MEAN	421.2	105.8	.00	.09	1.00	-.2	1.00	-.2			42.2	36.4	
P.SD	43.2	.4	.37	.00	.33	2.3	.33	2.4			7.5	.4	

Figure 22. Calibration 3, MSTAT-II instrument. Point measure correlation (column 10, ‘PTMEASUR’) indicates item polarity. Item statistics are sequenced in order of correlation. Item fit statistics are shown in columns 6 through 9 as standardsized (Zstd) and means square (Mnsq) indices of weighted infit and unweighted outfit.

Item STATISTICS: CORRELATION ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
22	442	106	-.19	.11	.80	-1.5	.80	-1.6	.56	.65	45.3	41.6	AT22
10	489	106	-.74	.11	1.42	2.7	1.40	2.7	.57	.63	35.8	43.1	AT10
8	341	106	.89	.10	1.35	2.5	1.33	2.4	.61	.67	39.6	39.3	AT8
13	455	106	-.34	.11	1.01	.1	1.03	.3	.65	.64	45.3	42.1	AT13
1	382	106	.46	.10	1.14	1.0	1.15	1.1	.66	.66	38.7	39.2	AT1
6	416	106	.09	.10	.92	-.5	.93	-.5	.67	.65	52.8	40.8	AT6
11	447	106	-.25	.11	.88	-.9	.85	-1.1	.67	.64	47.2	41.7	AT11
12	429	106	-.05	.10	.77	-1.8	.75	-2.0	.68	.65	51.9	41.3	AT12
16	413	106	.13	.10	.66	-2.9	.67	-2.8	.81	.65	55.7	40.8	AT16
MEAN	423.8	106.0	.00	.11	.99	-.1	.99	-.2			45.8	41.1	
P.SD	40.7	.0	.44	.00	.25	1.8	.24	1.8			6.4	1.2	

Figure 23. Calibration 4, Nine item instrument. Point measure correlation (column 10, 'PTMEASUR') indicates item polarity. Item statistics are sequenced in order of correlation. Item fit statistics are shown in columns 6 through 9 as standardsized (Zstd) and means square (Mnsq) indices of weighted infit and unweighted outfit.

Item STATISTICS: CORRELATION ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
22	336	106	-.21	.11	.88	-.9	.92	-.5	.56	.65	44.3	43.3	AT22
10	380	106	-.78	.12	1.40	2.7	1.39	2.5	.56	.62	37.7	45.6	AT10
8	249	106	.89	.12	1.31	2.2	1.28	2.0	.61	.67	39.6	41.8	AT8
13	348	106	-.36	.11	1.00	.0	1.17	1.3	.65	.64	45.3	44.4	AT13
11	340	106	-.26	.11	.90	-.8	.86	-1.1	.66	.64	48.1	43.7	AT11
1	281	106	.48	.11	1.12	.9	1.12	.9	.66	.67	42.5	41.7	AT1
12	324	106	-.06	.11	.85	-1.2	.91	-.7	.67	.65	51.9	42.7	AT12
6	310	106	.12	.11	.89	-.8	.89	-.8	.69	.66	56.6	42.8	AT6
16	306	106	.17	.11	.68	-2.8	.67	-2.8	.81	.66	56.6	42.8	AT16
MEAN	319.3	106.0	.00	.11	1.00	-.1	1.02	.1			47.0	43.2	
P.SD	36.4	.0	.46	.00	.22	1.6	.22	1.6			6.5	1.1	

Figure 24. Calibration 5, Nine item, five category instrument. Point measure correlation (column 10) indicates item polarity. Item statistics are sequenced in order of correlation. Item fit statistics are shown in columns 6 through 9 as standardsized (Zstd) and means square (Mnsq) indices of weighted infit and unweighted outfit.

Item STATISTICS: CORRELATION ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
22	336	106	-.34	.13	1.00	.1	1.07	.5	.62	.71	53.8	48.5	AT22
6	310	106	.08	.13	1.07	.5	1.07	.5	.71	.72	59.6	48.9	AT6
11	340	106	-.41	.13	.95	-.3	.91	-.6	.71	.70	59.6	48.3	AT11
1	281	106	.54	.13	1.18	1.3	1.21	1.5	.72	.74	52.9	47.3	AT1
16	306	106	.14	.13	.74	-2.1	.72	-2.2	.82	.73	64.4	48.8	AT16
MEAN	314.6	106.0	.00	.13	.99	-.1	1.00	.0			58.1	48.4	
P.SD	21.6	.0	.35	.00	.15	1.1	.17	1.3			4.2	.6	

Figure 25. Calibration 6, Five item, five category instrument. Point measure correlation (column 10, 'PTMEASUR') indicates item polarity. Item statistics are sequenced in order of correlation. Item fit statistics are shown in columns 6 through 9 as standardsized (Zstd) and means square (Mnsq) indices of weighted infit and unweighted outfit.

```
SUMMARY OF CATEGORY STRUCTURE.  Model="R"
-------------------------------------------------------------------
|CATEGORY   OBSERVED|OBSVD SAMPLE|INFIT OUTFIT|| ANDRICH |CATEGORY||
|LABEL SCORE COUNT %|AVRGE EXPECT|  MNSQ  MNSQ||THRESHOLD| MEASURE|
|-------------------+------------+------------++---------+--------|
|  1   1     105   4|  -.54  -.65|  1.27  1.49||  NONE   |( -2.88)| 1
|  2   2     263  11|  -.41  -.43|  1.07  1.17||   -1.46 |  -1.50 | 2
|  3   3     489  20|  -.26  -.22|   .85   .85||    -.94 |   -.69 | 3
|  4   4     606  24|  -.02   .00|   .76   .73||    -.33 |   -.04 | 4
|  5   5     567  23|   .25   .24|   .91   .92||     .18 |    .63 | 5
|  6   6     334  13|   .57   .55|   .96   .96||     .92 |   1.53 | 6
|  7   7     140   6|   .99   .99|  1.09  1.08||    1.63 |(  3.01)| 7
|-------------------+------------+------------++---------+--------|
|MISSING       4   0|   .11      |            ||         |        |
-------------------------------------------------------------------
OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

---------------------------------------------------------------------------------
|CATEGORY    STRUCTURE   |  SCORE-TO-MEASURE   | 50% CUM.| COHERENCE       |ESTIM|
| LABEL    MEASURE  S.E. | AT CAT. ----ZONE----|PROBABLTY| M->C C->M  RMSR |DISCR|
|------------------------+---------------------+---------+-----------------+-----|
|   1      NONE          |( -2.88) -INF   -2.20|         |   0%   0% 2.4744|     | 1
|   2        -1.46   .11 |  -1.50  -2.20  -1.05|   -1.86 |  38%  11% 1.6291|  .85| 2
|   3         -.94   .06 |   -.69  -1.05   -.36|    -.99 |  36%  39%  .9212|  .87| 3
|   4         -.33   .05 |   -.04   -.36    .28|    -.35 |  34%  62%  .6070| 1.05| 4
|   5          .18   .05 |    .63    .28   1.03|     .25 |  40%  43%  .9056| 1.14| 5
|   6          .92   .06 |   1.53   1.03   2.29|     .97 |  39%  22% 1.3938| 1.04| 6
|   7         1.63   .10 |(  3.01)  2.29  +INF |    1.98 |  63%   4% 1.9305| 1.00| 7
---------------------------------------------------------------------------------
M->C = Does Measure imply Category?
C->M = Does Category imply Measure?
```

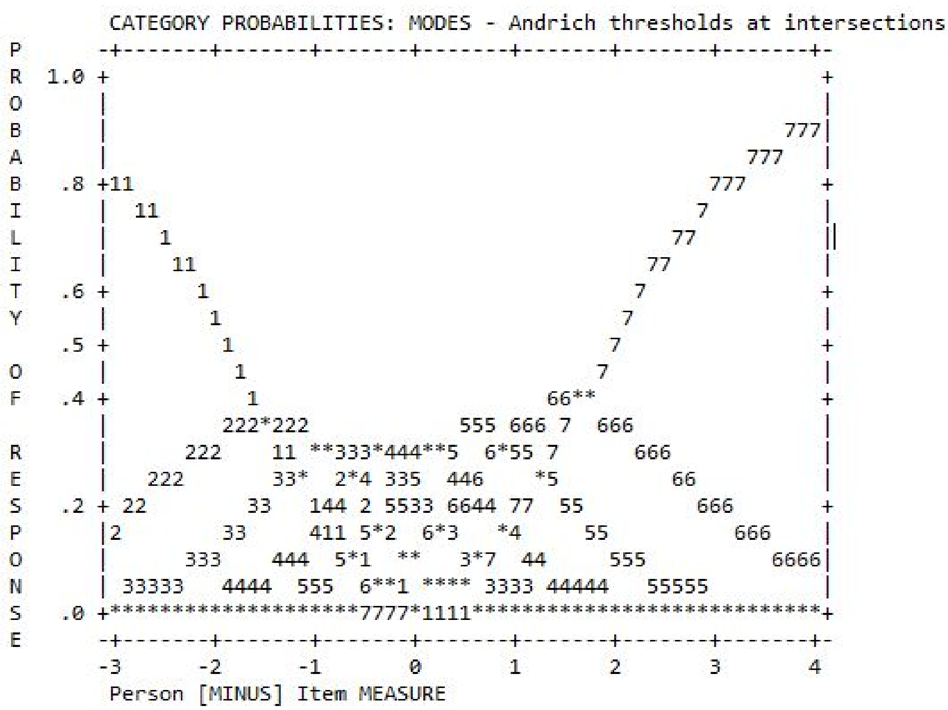

Figure 26. Calibration 1. MSTAT-I using seven categories.

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	ANDRICH THRESHOLD	CATEGORY MEASURE	
1	1	67	5	-.74	-.85	1.28	1.53	NONE	(-3.01)	1
2	2	159	11	-.60	-.58	1.03	1.05	-1.58	-1.63	2
3	3	308	21	-.35	-.32	.86	.86	-1.11	-.77	3
4	4	410	28	-.05	-.05	.72	.69	-.47	-.04	4
5	5	316	21	.24	.24	.93	.92	.35	.72	5
6	6	162	11	.59	.55	.97	.97	1.06	1.65	6
7	7	58	4	.89	.90	1.12	1.13	1.75	(3.13)	7
MISSING		2	0	.37						

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

CATEGORY LABEL	STRUCTURE MEASURE	S.E.	SCORE-TO-MEASURE AT CAT.	----ZONE----		50% CUM. PROBABLTY	COHERENCE M->C	C->M	RMSR	ESTIM DISCR	
1	NONE		(-3.01)	-INF	-2.33		0%	0%	2.3052		1
2	-1.58	.13	-1.63	-2.33	-1.16	-1.99	46%	17%	1.4595	.85	2
3	-1.11	.08	-.77	-1.16	-.40	-1.12	36%	42%	.8576	.95	3
4	-.47	.06	-.04	-.40	.33	-.42	41%	65%	.5864	1.07	4
5	.35	.06	.72	.33	1.14	.33	39%	42%	.9220	1.07	5
6	1.06	.08	1.65	1.14	2.41	1.10	39%	17%	1.4143	1.06	6
7	1.75	.14	(3.13)	2.41	+INF	2.10	0%	0%	2.0821	1.00	7

M->C = Does Measure imply Category?
C->M = Does Category imply Measure?

```
        CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+------+------+------+------+------+------+------+------+-
R  1.0 +                                                          +
O      |                                                          |
B      |111                                                     77|
A      |   111                                                777 |
B   .8 +      11                                            77    +
I      |        1                                          7      |
L      |         11                                      77       |
I      |           1                                    7         |
T   .6 +            1                                  7          +
Y      |             1                                7           |
    .5 +              1                              7            +
O      |               1                            7             |
F   .4 +                1                        667              +
       |              222*22  33  4444455555666 76666             |
R      |            22    13**  **3  554  6655 7     66           |
E      |          222     31  24   35   4*   7*5       66         |
S   .2 +        22       33  14422 5533  6 4 7   5       66       +
P      |     222       33    41   *    *6   *4    55        666   |
O      |2222         33    44  1*5 2266 3377  44    555        666|
N      |        33333  4444  555 1**622**73333  4444  5555        |
S   .0 +*************************77***11**************************+
E      -+------+------+------+------+------+------+------+------+-
       -4     -3     -2     -1      0      1      2      3      4
        Person [MINUS] Item MEASURE
```

Figure 27. Calibration 2. MSTAT-II using seven categories.

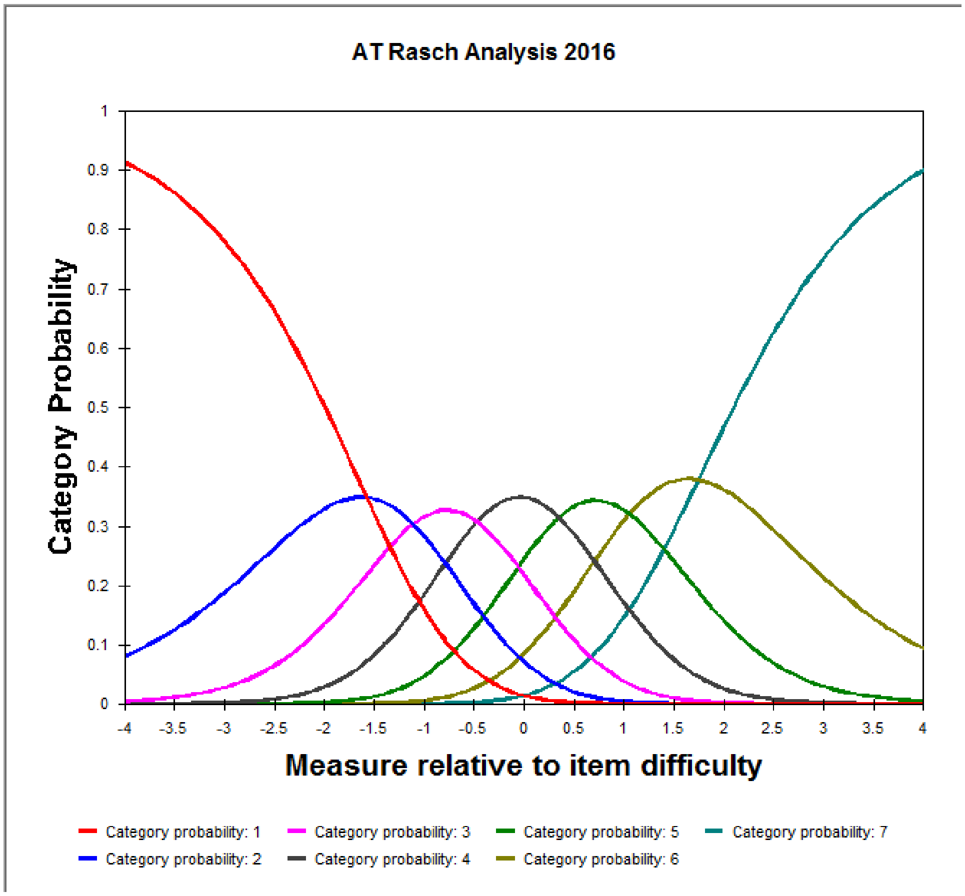

Figure 28. Calibration 2. MSTAT-II Category probability curves for 7 rating categories showing excessive overlap.

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	ANDRICH THRESHOLD	CATEGORY MEASURE	
1	1	42	3	-1.11	-1.09	1.01	1.02	NONE	(-3.52)	1
2	2	146	11	-.73	-.77	1.13	1.17	-2.17	-1.97	2
3	3	303	22	-.45	-.44	.95	.97	-1.33	-.95	3
4	4	399	29	-.09	-.09	.76	.74	-.54	-.08	4
5	5	310	23	.26	.28	1.03	1.03	.34	.85	5
6	6	146	11	.73	.68	1.00	1.00	1.23	2.02	6
7	7	30	2	1.08	1.10	1.04	1.03	2.47	(3.75)	7
MISSING		2	0	.40						

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

CATEGORY LABEL	STRUCTURE MEASURE	S.E.	SCORE-TO-MEASURE AT CAT.	----ZONE----		50% CUM. PROBABLTY	COHERENCE M->C	C->M	RMSR	ESTIM DISCR	
1	NONE		(-3.52)	-INF	-2.78		0%	0%	1.9770		1
2	-2.17	.16	-1.97	-2.78	-1.42	-2.48	50%	16%	1.4618	1.02	2
3	-1.33	.09	-.95	-1.42	-.51	-1.38	39%	44%	.8549	.96	3
4	-.54	.07	-.08	-.51	.37	-.52	43%	65%	.5716	.99	4
5	.34	.07	.85	.37	1.38	.36	42%	43%	.9239	1.01	5
6	1.23	.09	2.02	1.38	2.95	1.32	55%	20%	1.3716	1.07	6
7	2.47	.19	(3.75)	2.95	+INF	2.70	0%	0%	1.9405	1.00	7

M->C = Does Measure imply Category?
C->M = Does Category imply Measure?

```
        CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+------+------+------+------+------+------+------+------+-
R  1.0 +                                                         +
O      |                                                         |
B      |                                                         |
A      |11                                                       |
B   .8 +  11                                                   77+
I      |    11                                               77  |
L      |      1                                             7    |
I      |       1                                          77     |
T   .6 +        1                                        7       +
Y      |         11                                     7        |
    .5 +           1                                   7         +
O      |            1                             66666667       |
F   .4 +            2*2222                555  6        766      +
       |         222  1   2*3333344444*55    **5     77    66    |
R      |       22      1 33 2  4433  5 44  6   5  7        66    |
E      |     22         3*    **    *5    **     5*          66  |
S   .2 +   22          33  1  4  2  5 33  6  4    7 55         666+
P      |222           3     **     **    *6    4477    55        |
O      |         333    44  1155  2266 333  7744       55        |
N      |    33333   4444   555111666222 77**3   4444    55555    |
S   .0 +**********************************************************+
E      -+------+------+------+------+------+------+------+------+-
       -4     -3     -2     -1      0      1      2      3      4
        Person [MINUS] Item MEASURE
```

Figure 29. Calibration 3. MSTAT-II using seven categories and underfitting persons removed.

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	ANDRICH THRESHOLD	CATEGORY MEASURE	
1	1	27	3	-1.65	-1.69	1.06	1.10	NONE	(-4.03)	1
2	2	96	10	-1.28	-1.22	.94	.95	-2.72	-2.41	2
3	3	190	20	-.76	-.76	.98	.99	-1.67	-1.30	3
4	4	305	32	-.23	-.25	.77	.77	-.98	-.25	4
5	5	224	23	.33	.35	1.10	1.10	.35	.99	5
6	6	99	10	1.19	1.09	.99	.97	1.52	2.61	6
7	7	13	1	1.38	1.86	1.42	1.30	3.50	(4.70)	7

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

CATEGORY LABEL	STRUCTURE MEASURE	S.E.	SCORE-TO-MEASURE AT CAT.	----ZONE----		50% CUM. PROBABLTY	COHERENCE M->C	C->M	RMSR	ESTIM DISCR	
1	NONE		(-4.03)	-INF	-3.26		0%	0%	1.8261		1
2	-2.72	.21	-2.41	-3.26	-1.82	-2.99	43%	22%	1.2161	.98	2
3	-1.67	.11	-1.30	-1.82	-.80	-1.76	38%	47%	.8077	1.08	3
4	-.98	.08	-.25	-.80	.35	-.85	49%	64%	.5621	1.05	4
5	.35	.08	.99	.35	1.73	.34	44%	46%	.8839	.97	5
6	1.52	.12	2.61	1.73	3.79	1.64	65%	30%	1.1650	1.04	6
7	3.50	.30	(4.70)	3.79	+INF	3.62	0%	0%	1.8640	.86	7

M->C = Does Measure imply Category?
C->M = Does Category imply Measure?

```
         CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-
R  1.0 +                                                             +
O      |                                                             |
B      |11                                                           |
A      |  11                                                         |
B   .8 +    11                                                     77+
I      |      11                                                  7  |
L      |        1                                               77   |
I      |         1                                             7     |
T   .6 +          1                                           7      +
Y      |           1                             666666    7         |
    .5 +            1                          66      66 7          +
O      |             1 22          4444   55555 6        *6          |
F   .4 +            22*  22       44     4*5     *5       7  6       +
       |          22   1   2*333*3       5 4    6  5      77    66   |
R      |        2       1 33 2 4  3   5   4  6     55   7       6    |
E      |      22         *    *    335     4*       5 7          66  |
S   .2 +    22         33 1  4 22  553    66 4       *5            66+
P      |  22         33    *4    25    33 6    44   77  5            |
O      |222        33    44 11  5522   6*3     4*7       555         |
N      |       3333   444    5**1   ***   3337777 444        5555    |
S   .0 +*************************************************************+
E      -+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-
       -5    -4    -3    -2    -1     0     1     2     3     4     5
        Person [MINUS] Item MEASURE
```

Figure 30. Calibration 4 on 9 selected items from MSTAT-II using seven categories.

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	ANDRICH THRESHOLD	CATEGORY MEASURE	
1	1	123	13	-1.27	-1.24	1.00	1.02	NONE	(-2.78)	1
2	2	190	20	-.66	-.66	1.00	1.02	-1.38	-1.27	3
3	3	305	32	-.06	-.10	.76	.82	-.86	-.09	4
4	4	224	23	.53	.56	1.11	1.16	.52	1.24	5
5	5	112	12	1.57	1.57	1.05	1.09	1.71	(3.00)	6

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

CATEGORY LABEL	STRUCTURE MEASURE	S.E.	SCORE-TO-MEASURE AT CAT.	----ZONE----		50% CUM. PROBABLTY	COHERENCE M->C	C->M	RMSR	ESTIM DISCR	
1	NONE		(-2.78)	-INF	-2.07		80%	16%	1.2253		1
2	-1.38	.11	-1.27	-2.07	-.67	-1.75	38%	51%	.7875	1.07	3
3	-.86	.08	-.09	-.67	.53	-.71	49%	65%	.5272	1.05	4
4	.52	.08	1.24	.53	2.19	.51	44%	46%	.8639	.92	5
5	1.71	.12	(3.00)	2.19	+INF	1.94	73%	29%	1.1380	1.01	6

M->C = Does Measure imply Category?
C->M = Does Category imply Measure?

```
        CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+---------+---------+---------+---------+---------+---------+-
R  1.0 +                                                             +
O      |                                                             |
B      |                                                             |
A      |                                                             |
B   .8 +111                                                          +
I      |   11                                                     555|
L      |     11                                                 55   |
I      |       11                                             55     |
T   .6 +         11                                         55       +
Y      |           1                                       5         |
    .5 +            11                                   55          +
O      |              1          3333333     444444444*5             |
F   .4 +               11       333       33*44        5 444         +
       |            222222*222**2         44 333     55     444      |
R      |         222       1*3   22     44      33 55         444    |
E      |      222         33 1     22244          *3            444  |
S   .2 + 2222            33    11    4422        55  33            4+
P      |2            333        1*44     222  555      333           |
O      |          333        444 111     5**2             3333       |
N      | 3333333      444444      5*****     222222          3333333 |
S   .0 +**************555555555555     1111111111*****************+
E      -+---------+---------+---------+---------+---------+---------+-
       -3        -2        -1         0         1         2         3
        Person [MINUS] Item MEASURE
```

Figure 31. Calibration 5 on 9 selected items from MSTAT-II using five categories.

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	ANDRICH THRESHOLD	CATEGORY MEASURE	
1	1	65	12	-1.72	-1.73	1.05	1.07	NONE	(-3.19)	1
2	2	103	19	-1.05	-.95	.91	.96	-1.80	-1.60	3
3	3	200	38	-.17	-.20	.65	.62	-1.24	-.14	4
4	4	108	20	.78	.66	1.07	1.12	.82	1.58	5
5	5	54	10	1.85	2.08	1.28	1.23	2.22	(3.47)	6

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

CATEGORY LABEL	STRUCTURE MEASURE	S.E.	SCORE-TO-MEASURE AT CAT.	----ZONE----		50% CUM. PROBABLTY	COHERENCE M->C	C->M	RMSR	ESTIM DISCR	
1	NONE		(-3.19)	-INF	-2.46		89%	25%	1.1338		1
2	-1.80	.16	-1.60	-2.46	-.90	-2.15	47%	64%	.6911	.98	3
3	-1.24	.12	-.14	-.90	.71	-1.01	62%	78%	.4395	1.13	4
4	.82	.12	1.58	.71	2.63	.75	56%	44%	.8304	1.04	5
5	2.22	.20	(3.47)	2.63	+INF	2.41	67%	36%	1.1621	.79	6

M->C = Does Measure imply Category?
C->M = Does Category imply Measure?

```
        CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+------+------+------+------+------+------+------+------+-
R  1.0 +                                                        +
O      |                                                        |
B      |11                                                      |
A      |  11                                                  55|
B   .8 +    11                                              55  +
I      |      11                                          55    |
L      |        1                                        5      |
I      |         11                                    55       |
T   .6 +           1                                  5         +
Y      |            1           33333                5          |
    .5 +             1        33     33     44444   5           +
O      |              1     33        33444      44*            |
F   .4 +               *222 3          43       55 44           +
       |            222 1  *22        44  33   5    44          |
R      |          22    *3   22      4     3  5       44        |
E      |        22     3 1     22  44       3*          44      |
S   .2 +      222     3   11     **        55 3           44    +
P      |  222        33     1  44  22    55    33           44  |
O      |22        333        **1    22255        333            |
N      |      3333      44444   111*55552222          33333     |
S   .0 +****************5555555555 11111111********************+
E      -+------+------+------+------+------+------+------+------+-
       -4     -3     -2     -1      0      1      2      3      4
        Person [MINUS] Item MEASURE
```

Figure 32. Calibration 6 on 5 items of *general ambiguity* using five categories.

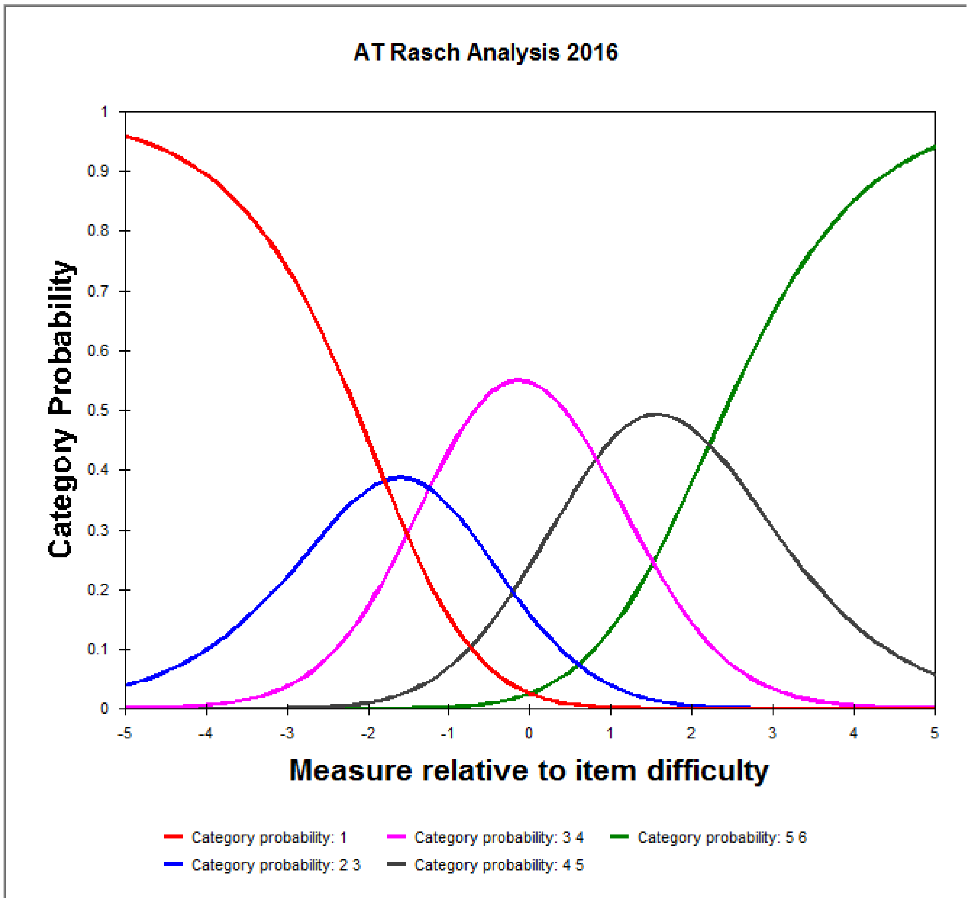

Figure 33. Calibration 6. Category probability curves for 5-rating categories on 5-Item subset of MSTAT-II (General Ambiguity) showing overlap.