

Practical Assessment, Research, and Evaluation

Volume 12 *Volume 12, 2007*

Article 14

2007

Assessing unidimensionality: A comparison of Rasch Modeling, Parallel Analysis, and TETRAD

Chong Ho Yu

Sharon Osborn-Popp

Samuel DiGangi

Angel Jannasch-Pennell

Follow this and additional works at: <https://scholarworks.umass.edu/pare>

Recommended Citation

Yu, Chong Ho; Osborn-Popp, Sharon; DiGangi, Samuel; and Jannasch-Pennell, Angel (2007) "Assessing unidimensionality: A comparison of Rasch Modeling, Parallel Analysis, and TETRAD," *Practical Assessment, Research, and Evaluation*: Vol. 12 , Article 14.

DOI: <https://doi.org/10.7275/q7g0-vt50>

Available at: <https://scholarworks.umass.edu/pare/vol12/iss1/14>

This Article is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Practical Assessment, Research, and Evaluation by an authorized editor of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Copyright is retained by the first or sole author, who grants right of first publication to the *Practical Assessment, Research & Evaluation*. Permission is granted to distribute this article for nonprofit, educational purposes if it is copied in its entirety and the journal is credited.

Volume 12, Number 14, October 2007

ISSN 1531-7714

Assessing unidimensionality: A comparison of Rasch Modeling, Parallel Analysis, and TETRAD

*Chong Ho Yu, Sharon Osborn Popp, Samuel DiGangi, Angel Jannasch-Pennell,
Arizona State University*

The evaluation of assessment dimensionality is a necessary stage in the gathering of evidence to support the validity of interpretations based on a total score, particularly when assessment development and analysis are conducted within an item response theory (IRT) framework. In this study, we employ polytomous item responses to compare two methods that have received increased attention in recent years (Rasch model and Parallel analysis) with a method for evaluating assessment structure that is less well-known in the educational measurement community (TETRAD). The three methods were all found to be reasonably effective. Parallel Analysis successfully identified the correct number of factors and while the Rasch approach did not show the item misfit that would indicate deviation from clear unidimensionality, the pattern of residuals did seem to indicate the presence of correlated, yet distinct, factors. TETRAD successfully confirmed one dimension in the single-construct data set and was able to confirm two dimensions in the combined data set, yet excluded one item from each cluster, for no obvious reasons. The outcomes of all three approaches substantiate the conviction that the assessment of dimensionality requires a good deal of judgment.

The evaluation of assessment dimensionality is a necessary stage in the gathering of evidence to support the validity of interpretations based on a total score, particularly when assessment development and analysis is conducted within an item response theory (IRT) framework. Unidimensionality refers to the existence of one underlying measurement construct (dimension) that accounts for variation in examinee responses. Violating this assumption could severely bias item and ability parameter estimation. In this study, we employ polytomous item responses to compare two methods that have received increased attention in recent years: Rasch model analysis (Rasch, 1960/1980) and Parallel analysis (Horn, 1965) with TETRAD (Glymour, 1982), a method for evaluating assessment structure that is less well-known in the educational measurement community, though confirmatory TETRAD analysis has been developed by Bollen and Ting (1993, 1998, 2000) for identifying causal indicators.

Many methods of investigating unidimensionality are available. Two notable reviews of methods and indices of unidimensionality have been conducted within the last

twenty-five years. Hattie's (1985) review evaluated numerous standard approaches and showed that many lacked empirical support for the adequate assessment of unidimensionality. More recently, Tate (2003) conducted a review of methods and indices employed with dichotomous items, finding that options for assessing dimensionality had expanded and improved, and that most methods perform effectively "within the limits of their associated perspectives and assumptions." While this study highlights the TETRAD method, compared to Rasch model and Parallel analyses, there are several established methods that are well-documented and widely used. Test of Essential Dimensionality (DIMTEST) (Stout, 1987, 1990), Dimensionality Evaluation To Enumerate Contributing Traits (DETECT) (Kim, 1994; Zhang & Stout, 1999), and Hierarchical Cluster Analysis with Proximity Matrix (HCA/CCPROX) (Roussos, Stout, & Marden, 1998) are well-established examples, which are nonparametric in essence and can operate in either exploratory or confirmatory mode¹. Other methods widely employed to assess dimensionality include confirmatory factor analysis through structural equation modeling (SEM), applied using programs

like EQS (Bentler & Wu, 1993), LISREL (Jöreskog & Sörbom, 1989) and Mplus (Muthén & Muthén, 1998).

Although factor analysis has been widely used for evaluating dimensionality, factor analysis often performs inadequately by confounding variation in item difficulty with dimensionality. As a result, the true number of latent factors is often over-estimated (Stout, Nandakumar, & Habing, 1996). In addition, factor analysis based on the Pearson correlation matrix has been regarded as problematic when applied to dichotomous tests due to nonlinearity. In an attempt to rectify the situation, the replacement of the Pearson matrix by the tetrachoric matrix (based on correlations which estimate response distributions for the underlying continuous variables assumed to be represented by the observed dichotomous variables) has been suggested. However, using the tetrachoric matrix requires very intense computing resources and the performance gain is not always significant (Meara, Robin & Sireci, 2000). Other concerns regarding the use of the tetrachoric matrix for factor analysis have also been raised. Using a tetrachoric matrix for factor analysis may fail to produce just one common factor unless certain normality assumptions are met (Lord & Novick, 1968). Tetrachorics can also be severely affected by guessing behavior on multiple-choice exams with difficult items (Carroll, 1945; Lord, 1980). Indeed, when the test has a wide range of difficulties, tetrachorics are not considered dependable unless the sample size is larger than 2,000 (Roznowski, Tucker, & Humphreys, 1991). However, some psychometricians have contended that the core issue of dimensionality is the nonlinear function between the fitted factor scores, not the item difficulties per se. Thus, nonlinear factor analysis was proposed as a remedy (McDonald, 1967, 1981, 1985, 1997, 1999). A variant of nonlinear factor analysis is full-information factor analysis, which is said to improve upon conventional factor analysis by jointly estimating multiple thresholds and factor loadings for each item (Bock, Gibbons, & Muraki, 1988). It was developed for binary responses, but Gibbons, et al. (2007) expressed concern that there has been little progress in multidimensional extensions of full information factor analysis for graded response models due to the added computational complexity.

Despite demonstrations of the basic equivalence of IRT and factor analysis models, (see e.g., Takane & de Leeuw, 1987), some researchers have preferred IRT models over factor analysis and principal component analysis for examining dimensionality. In the 1960s, Wright and Panchapakesan (1969) asserted that if a given set of items fit the Rasch model (a class of IRT models also referred to as one-parameter logistic), then there is evidence that these items refer to a unidimensional construct. In recent years, some researchers went even further to suggest that the Rasch model is a superior method to factor analysis in terms of confirming a factor structure (Waugh & Chapman, 2005). When misfits are identified and removed from the scale, unidimensionality can be preserved during the Rasch diagnosis. It seems redundant

to check the factor structure first and then run a Rasch analysis later. However, the Rasch model has been found to not be as effective in situations with uncorrelated factors (Smith, 1996; Wright, 1997). A study by Tennant and Pallant (2006) using simulated scales consisting of polytomous items indicated that the Rasch model fit statistics performed poorly when two factors with almost equal number of items were interrelated. Misfit diagnosis as a means of dimensionality diagnosis works well if there is a dominant factor with a much larger number of items and a few misfits that emerge relative to the dominant dimension. Smith and Miao (1994) suggest that principal component analysis and Rasch measurement may be used to complement each other, “assuring the widest possible coverage of different combinations of common variance and proportion of items loading on the second factor” (p. 327). It has also been suggested that exploratory factor analysis (EFA) with parallel analysis (PA) may be used prior to the application of IRT models in order to give early indications of any dimensionality issues (Budescu, Cohen, & Ben-Simon, 1997; Weng & Cheng, 2005).

In the arena of EFA, more and more researchers are skeptical of conventional criteria for extracting factors, such as the Kaiser criterion (Zwick & Velicer, 1986; Velicer, Eaton, & Fava, 2000). As a remedy, many factor modelers prefer parallel analysis (PA) to the Kaiser criterion (Horn, 1965; Hayton, Allen, Scarpello, 2004; Weng & Cheng, 2005). The logic of parallel analysis resembles that of re-sampling in the sense that the number of factors extracted should have eigenvalues greater than those in a random matrix. To materialize this theoretical notion, the algorithm generates a set of random data correlation matrices by bootstrapping the data set, and then the average eigenvalues are computed. Next, the observed eigenvalues are compared against the re-sampled eigenvalues, and only factors with observed eigenvalues greater than those from re-sampling are retained.

Some researchers suggest that exploratory tetrad analysis (ETA) outperforms factor analysis in removing impure indicators that do not belong to the target factor (Glymour, Scheines, Spirtes, & Kelly, 2000). Factor analysis and Rasch modeling are widely used by psychological and educational researchers, however, TETRA Difference (TETRAD) analysis (Scheines, Glymour & Spirtes, 2005) is less well-known in the psychological and educational research community in spite of its theoretical soundness. The objective of this article is to compare the efficacy of Rasch misfit diagnosis as dimensionality detection implemented in RUMM (Andrich, Lyne, Sheridan, & Luo, 1997), factor analysis using PA implemented in VisTa (Ledesma, & Valero-Mora, 2007; Young, 1999), and TETRAD analysis. A practical demonstration of the three approaches was conducted, using responses from instruments employed in the assessment of a professional development program in mathematics education.

WHAT IS TETRAD?

Unlike other SEM software applications that are usually programmed by psychologists or statisticians, TETRAD was conceptualized by a prominent philosopher of science, Clark Glymour (1982), and was developed by a group of researchers in the philosophy department at Carnegie Mellon University (CMU) (Glymour, et al., 2000). In order to pierce to the core of TETRAD, it is important to introduce the historical roots and philosophical foundations on which TETRAD is built. In this section, a simple example given by the CMU group will be used to illustrate how TETRAD can be applied to “purify” a measurement model. Next, the focus will be shifted to key assumptions, premises, and characteristics of TETRAD.

To illustrate the basic idea of TETRAD, a factor model measuring the common features of democratic societies proposed by Bollen (1980) will be used as an example. According to Bollen, a democratic society, in theory, might be characterized by the following indicators: Press freedom (PF), freedom of group opposition (FG), (lack of) government sanctions (GS), fairness of elections (FE), executive selection (ES), and legislative selection (LS). Since Bollen attempted to formulate a unidimensional factor model, all of the preceding variables were supposed to be loaded into a single theoretical construct (I). But there is a correlation between FG and GS, between ES and LS, as well as between FG and LS. Whether this latent construct *T* is a cause for those diverse political behaviors or is just a summarization of observed variables is an ongoing debate.

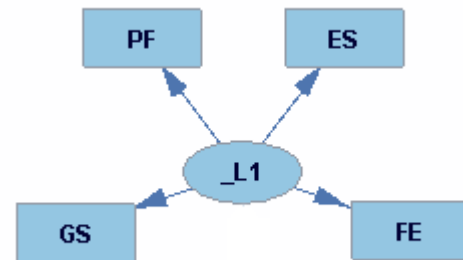
Since the latent factor *T* has been specified in the model and the objective of this search is to examine whether the measured political behaviors can be explained by a common theme, the “purify” procedure is employed in this example. According to Scheines, Sprites, Glymour, Meek and Richardson (1998), the purify algorithm discovered that including FG and LS, as suggested by Bollen’s initial model, would lead to a violation of unidimensionality. The best model recommended by TETRAD contains a latent construct *_L1*, and is shown in Figure 1. In other words, the degree of democracy of a society can be sufficiently indicated by PF, GS, ES, and FE.

This simple example, in which TETRAD yields a single “best” model, is chosen just for the sake of clarity. It is important to note that usually TETRAD outputs a family of models rather than a single best solution. The CMU group suggests that the proposed cluster of models can be further tested by other SEM software applications, such as EQS and LISREL. Hence, the automated path generation is an aid to, but not a replacement of, subsequent testing by human researchers.

The search algorithm in TETRAD, as its name implies, utilizes Spearman’s tetrad difference equations or vanishing tetrads (Hart & Spearman, 1913), and thus, in order for the

program to find a subset of measured variables for a factor model as indicated in this example, at least four indicators (measured variables) per factor is required. Tetrad refers to the

Figure 1. SEM search result



difference between the product of a pair of covariances and the product of another pair among four random variables. For example, if there are four indicators, there will be three tetrad difference equations:

$$D1 = \sigma_{12} \sigma_{34} - \sigma_{13} \sigma_{24}$$

$$D2 = \sigma_{13} \sigma_{24} - \sigma_{14} \sigma_{23}$$

$$D3 = \sigma_{14} \sigma_{23} - \sigma_{12} \sigma_{34}$$

where *D* = Tetrad difference and *Sigma* = variance

If the tetrads are zero, they are called vanishing tetrads, which indicate that the four variables share a common latent factor. In other words, the researcher should obtain zero partial correlations when the model is *linear*. In TETRAD, significance tests are conducted on partial correlations to determine whether two variables are independent given fixed values for some set of other variables. This requirement is called conditional independence, which will be discussed in a later section.

Although the tetrad difference equation was the first approach in attempts to detect latent constructs, it was eventually over-shadowed by other techniques such as principal components (Hotelling, 1933), maximum likelihood (Lawley & Maxwell, 1971) and weighted least squares (Browne, 1984). Nonetheless, after the vanishing tetrad approach was revived by Glymour and his colleagues in recent years, many researchers also endorsed it in various applications. For example, when Mulaik and Millsap (2000) defended the use of four indicators per factor in their four-step approach for testing a SEM, they praised the tetrad approach for its merits of over-determining the latent variable. To be specific, one can always find a perfect fit between a unidimensional factor model with three positively correlated indicators. In this case no test of the single-factor model is possible with this set up. However, four positively correlated variables may not have a single common factor, and therefore,

this over-identified common-factor model is testable or refutable.

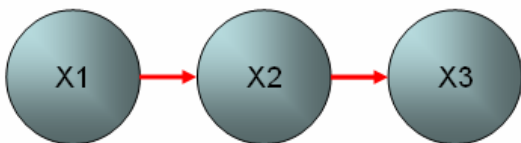
Besides reviving the concept of Spearman's TETRA Difference, the CMU group also introduces new assumptions into the new TETRAD approach, such as Causal Markov Condition (CMC) and Faithfulness assumption (FA). The TETRAD approach is essentially a causal discovery methodology. Instead of just conveniently reducing a large number of measured items into one construct for the sake of manageability, TETRAD modelers believe that the latent construct and the observed items are causally related, and therefore both CMC and FA are introduced to facilitate causal discovery. Like EQS, LISREL, and AMOS, TETRAD is capable of both factor modeling and structural equation modeling. But according to the CMU group, if there is no causal structure in the factor model, it makes no sense to make causal inferences in the structural model at all. Although this notion is philosophically fascinating and thus it will not do any justice to TETRAD without mentioning it, it is not the focal point of this article. In the subsequent sections CMC and FA will be illustrated in order to present a complete picture of TETRAD, nonetheless, the final section will concentrate on the efficacy of TETRAD in terms of examining dimensionality.

ASSUMPTIONS OF TETRAD

Causal Markov Condition

In a causal model, the joint probability distribution over the variables must satisfy the Causal Markov Condition (CMC) (Druzdzel & Glymour, 1995). Let G be a causal graph, in which variables in a set called V are represented by vertices or nodes (circles) and cause-effect relationships are denoted by directional arrows (see Figure 2). Let P be an associated probability distribution over V . In G , $(X_1 \rightarrow X_2 \rightarrow X_3)$ means X_1 causes X_2 , and X_2 causes X_3 .

Figure 2. Example of the Causal Markov Condition



Intuitively, the direct descendent of a variable is its effect and the immediate parent of a variable is its cause. Obviously, X_1 is the immediate parent of X_2 , X_3 is the direct descendent of X_2 , and X_3 is a non-descendent of X_1 . The configuration of $(X_1 \rightarrow X_2 \rightarrow X_3)$ is called a "collider complex." CMC requires that conditional on its parent (its direct cause in V) each variable is probabilistically independent from non-descendants, which include all other variables except its

immediate effect. In the causal graph G , X_3 is conditionally independent from X_1 given X_2 if $\text{Prob}[X_3|X_2, X_1] = \text{Prob}[X_3|X_2]$. In philosophy terminology, X_2 is said to "screen off" the correlation between X_1 and X_3 when X_1 and X_3 are considered uncorrelated because the presence of X_1 does not increase the probability of X_3 given the "screener" X_2 (Reichenbach, 1956).

Hence, CMC is an assumption of the path model in which relationships among variables are structured without a feedback loop. In other words, TETRAD assumes that the causal structure is acyclic. In this example, CMC accepts a collider complex only. If X_3 is said to be a cause of X_1 , such as $X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_1$, then X_2 will cease to be a screener and thus CMC is violated. The TETRAD group realizes that on some occasions this assumption may be unrealistic. For instance, in a study regarding student retention in US colleges using the TETRAD approach, most of the variables under study might influence the image of the university. The image, in turn, might influence all other variables. Nevertheless, the researchers argue that the acyclicity assumption is acceptable because all feedback processes in this case are extremely slow acting, in the sense that it takes years or even decades for the feedback cycle to happen (Druzdzel & Glymour, 1995). Although in SEM specific methodologies had developed to tackle models with feedback loops, which are also known as non-recursive models, it is generally agreed that it is difficult to determine whether a non-recursive model is identified (Kline, 2006).

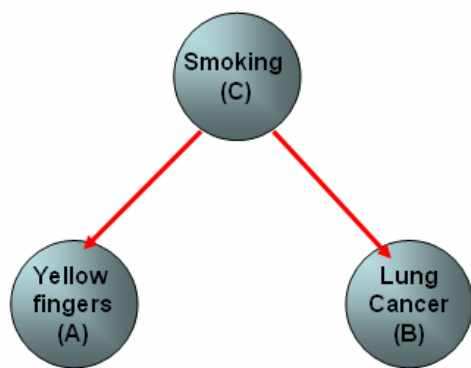
An example in ecology can illustrate CMC. In studying transitions of vegetation, ecologists realize that a location occupied by a species S_1 at time t will be replaced by species S_2 at time $t+1$ (Shipley, 2000). This is considered a Markovian process in the sense that changes in the vegetation at $t+1$ depend on the state of the vegetation at t , but not the distant past, such as $t-1$, $t-2$, and $t-3$. Simply put, in causal modeling once the researcher knows the direct cause of an event, then knowledge of indirect causes does not provide additional information. Glymour (2001) was critical of sequential search procedures because in a stepwise process each variable is selected or deselected by conditioning one variable on all others. On the contrary, the search algorithm based on CMC treats relationships among variables as conditionally independent from the indirect causes.

CMC also implies the common cause principle proposed by Reichenbach (1956) and advocated by Glymour and his colleagues (Glymour, 1982; Glymour, et al., 2000). According to the common cause principle, if a system of variables satisfies the Markov Condition, and they have a high degree of association, then there does exist a latent construct (factor) causing them. The common cause principle is the underlying assumption of the factor model. Hence, TETRAD is said to be a tool for constructing a one-dimensional factor model.

Reichenbach’s (1956) common cause principle (CCP) is very important to factor modeling. CCP is subsumed by the CMC. In a similar vein to CMC, CCP states that simultaneous correlated events that are not causally linked must have a prior common cause that acts as a screener. When A and B are correlated, but in actuality C is the common cause of A and B, C is said to screen off the pseudo causal relations among A and B, given that A and B are independent conditioning on C. The common cause principle can be effectively applied in factor analysis to inferences of the existence of a latent cause (Glymour, et al., 2000).

The relationships among A, B, and C can be expressed in the following fashion: If $P(A|B) > P(A)$, apparently the presence of B increases the probability of A. Conversely, if $P(B|A) > P(B)$, the presence of A seems to be a contributor to the higher probability of B. But if $P(B|A \& C) = P(B|C)$, then C is said to screen A off from B. In other words, C renders A probabilistically irrelevant to B. For example, there is an apparent association between yellow-stained fingers (A) and lung cancer (B). However, it is absurd to think that whitening one’s finger nails could reduce the risk of suffering from lung cancer. It is believed that both lung cancer and yellow-stained fingers have a common cause: smoking (C). According to CCP, smoking (C) screens yellow-stained fingers (A) off from lung cancer (B). The preceding relationships among A, B, and C form a conjunctive fork (Reichenbach, 1956), as shown in Figure 3.

Figure 3. Conjunctive fork of smoking, yellow-stained fingers and lung cancer.



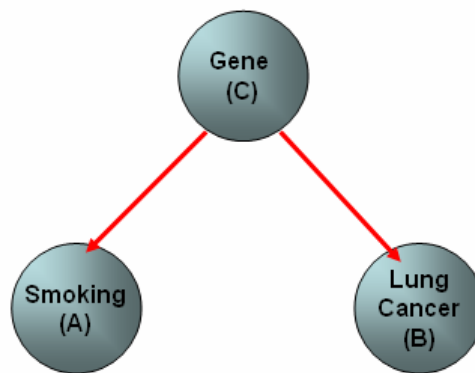
These probability relationships characterize a conjunctive fork:

- (1) $P(A \& B) > P(A)P(B)$
- (2) $P(A \& B | C) = P(A | C)P(B | C)$
- (3) $P(A \& B | \sim C) = P(A | \sim C)P(B | \sim C)$
- (4) $P(A | C) > P(A | \sim C)$
- (5) $P(B | C) > P(B | \sim C)$

Take smoking as an example again. Under Condition (1), the probability of having lung cancer and nicotine stains on fingers together is higher than the product of the probabilities of these two separate events. Yellow nails can be a result of the “yellow nail syndrome.” Lymphedema, especially of the ankles, and compromised respiration may be the cause. On the other hand, lung cancer can be caused by second-hand smoking. High levels of pollution, radiation, and asbestos exposure may also lead to lung cancer. Conditions (2) and (3) state that lung cancer and yellow fingers are conditionally independent when the common cause, smoking, is present or absent. Conditions (2) and (3) imply that both C and “not C” screen off A from B. Conditions (4) and (5) state that yellow nails and lung cancer are more probable, conditional on smoking. Conditions (2) through (5) entail (1).

However, keep in mind that CCP is a principle that points to the presence of some screener, but it doesn’t by itself say what the screener is. In the previous example, smoking is said to be a common cause of both yellow-stained fingers and lung cancer and these relationships form a conjunctive fork. But the conjunctive fork shown in Figure 4, which also satisfies CCP, states that smoking is not a cause of lung cancer. Rather, it is said that both lung cancer and smoking are caused by a specific genetic configuration. Needless to say, Figures 3 and 4 contradict each other.

Figure 4. Conjunctive fork of gene, smoking, and lung cancer.

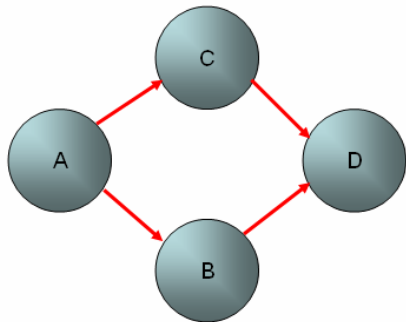


Faithfulness assumption

It is important to point out that even if vertices in a causal graph are independent, it does not necessarily mean that this independence must be entailed by CMC. For example, let P be a probability distribution on a causal graph named G consisting of four vertices, namely, A, B, C, and D (Figure 5). In linear models, independence can arise if the product of the partial regression coefficients for D on C and C on A cancels the corresponding product of D on B and B on A. However, if this canceling out effect is denied and the conditional independence relations true in P are entailed by CMC applied

to G, then P and G are treated as being faithful to each other. In this case, G is considered a perfect map of P (Spirtes, Glymour, & Scheines, 2000). This additional assumption to CMC is called the faithfulness assumption (FA).

Figure 5. Causal graph consisting of four vertices.



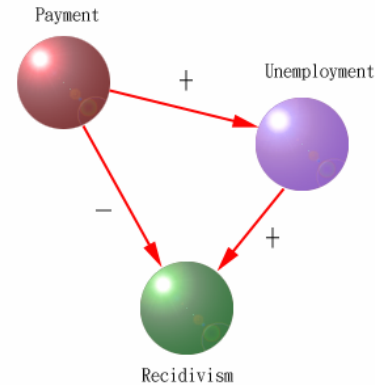
According to FA, statistical constraints arise from structure, not coincidence. As the name implies, FA supposes that probabilistic dependencies will faithfully reveal causal connections and there are no causes that are independent of effects. In other words, all independence and conditional independence relations among observed variables are consequences of the CMC applied to the true causal structure. For example, a research study (cited in Glymour et al., 2000) indicates that providing financial aid to released prisoners did not reduce recidivism. An alternate explanation is that free money discourages employment, and unemployment has a positive effect on recidivism, while financial aid tends to lower recidivism. As a result, these two effects cancel out each other (Figure 6). However, this explanation violates the faithfulness assumption. In other words, FA would rule out the particular values coincidentally canceling each other out in the model of Figure 6.

In the beginning of this section, Bollen’s model of democracy was discussed. Taking Bollen’s assertion about indicators of democracy into account, Li and Reuveny (2003) explore the cause and effect relationships between globalization and democracy. As you can expect, the answer is not dichotomous. One cannot easily side with conservatives to assert that globalization promotes democracy; likewise, one also cannot concur with leftists to accuse free trade of promoting oppression. Li and Reuveny identify four intermediate effects between globalization and democracy, namely, trade openness, portfolio investment inflows, foreign direct investment inflows, and the spread of democratic ideas across countries. It was found that trade openness and portfolio investment inflows negatively affect democracy

Foreign direct investment inflows positively affect democracy, but the effect weakens over time. The spread of democratic ideas promotes democracy persistently over time (see Figure 7). The significant point is that in this example, in

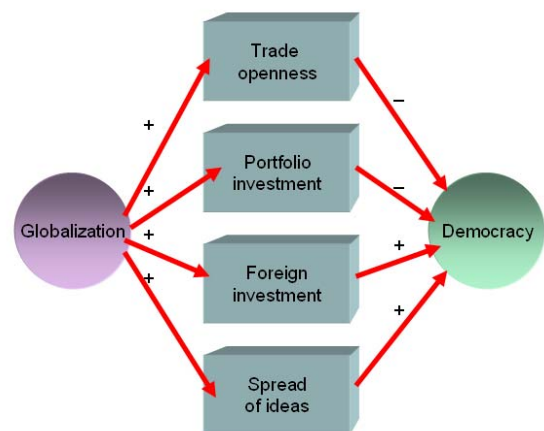
which two results created by globalization promote democracy and two others do the opposite, Li and Reuveny did not reject

Figure 6. Example of the Faithfulness assumption.



the causal link between globalization and democracy. Instead, Li and Reuveny gave concrete recommendations to policymakers based on the preceding different causal paths. In short, FA rules out that some effects cancel out each other and thus no causal effect can be discovered. In addition, it is a common pattern to trace the “root cause” in political debate and usually proposed actions to address the immediate causes are dismissed as lacking insight. Some critics of globalization asserted that globalization is a product of neoliberalism or market fundamentalism, which is considered a flawed ideology (Stiglitz, 2002). However, according to CMC, once the researcher knows the direct cause of an event, then knowledge of indirect causes does not provide additional information. In other words, information of the relationships between globalization, trade openness, portfolio investment, foreign investment, spread of policy ideas, and development of democracy is sufficient for policy advice.

Figure 7. Positive and negative intermediate effects between globalization and democracy



METHOD

Data

The data used in this study consists of posttest rating scale responses from two short self-report instruments, collected from 135 Grade 5 through 7 teachers participating in a series of professional development workshops in mathematics education from 2004 to 2006. In addition to other required assessments, participants were asked to rate their understanding of mathematics concept areas before and after workshop participation on seven general concept areas in geometry and measurement and on five general concept areas in data analysis and probability. A visual, eleven-point rating scale accompanied each concept area (e.g., “Identify and recognize the relationships between parts of a circle”), with anchors indicating no understanding (0 points), complete understanding (10 points), and all other scale points indicating approximate percentage of understanding (e.g., a 5 indicates “I understand about 50% of the concepts and their application to solving problems”). Posttest self-report ratings were collected in two combined posttest and retrospective pretest self-report instruments, with one containing the geometry and measurement concepts (G-items) and one containing the data analysis and probability items (D-items). The rationale of using actual data rather than simulated data is that in the latter usually large sample size, extreme cases and rare distributions are generated, but the applicability of methods confirming factor structure in realistic settings is the focus here. As mentioned before, in some simulation studies at least 2,000 subjects are needed in order to yield usable information from factoring tetrachorics. But as a practical illustration, this study aims to provide informative guides to test developers who are often confined to use small data sets.

Procedure

Data were organized into two data sets. The seven G-items were analyzed as one scale to assess the presence of a single factor (one-factor data set). The G-items and D-items were then combined to assess how the approaches perform in detecting the presence of two factors (two-factor data set).

Basic exploratory data analysis and data visualization were implemented in JMP (SAS Institute, 2007) and DataDesk (DataDescription Inc., 2007) to examine the data structure, including their distributions, potential outliers, and inter-relationships among variables. TETRAD is said to work best with normally distributed data. Thus, the distributions of the item responses are examined by normal quantile plots. Item inter-relationships were examined with scatterplot matrices.

The Rasch ordered-category, or “partial-credit,” rating scale model (Andrich, 1978; Masters, 1982) was applied to the data sets, using RUMM (Rasch Unidimensional Measurement Models) software (Andrich, Lyne, Sheridan, & Luo, 1997).

These analyses provided estimates of response thresholds among the scale points, i.e., the point at which the probability of choosing one scale point becomes higher than another. The probability of choosing a higher scale point should increase as the participant’s perception of their conceptual understanding increases. In the case of attitude or self-reporting rating scales, where response options have been developed specifically to reflect increasing or decreasing views or perceptions, thresholds that do not progress along the intended response continuum, are problematic. Disordered or proximal threshold estimates may indicate that the scale does not fit the structure of response content well.² Threshold estimates were reviewed for disorder, and then item residuals were examined and plotted for visual inspection to assess item fit and dimensionality for each data set.

Exploratory factor analyses, using parallel analysis (PA), were conducted on each data set using VisTa software. Parallel scree plots, eigenvalues, and eigenvalues generated from the set of random data correlation matrices produced in the analyses were examined to assess dimensionality of the two sets of responses. The number of factors extracted should have eigenvalues greater than those generated from the random matrix to reflect the dimensionality for each data set. In addition, because PA simulates a random data matrix, one would never see the same re-sampled eigenvalues twice. In order to achieve stability, the number of samples entered should be larger than the default (100). The number of samples generated for the PA analyses in this study was 200 (see Figure 8).

Figure 8. Options for PA

TETRAD analyses were then applied to each dataset. Like CFA, a researcher may specify which items are loaded into which factors when using TETRAD. The G-items were specified as a single factor in an analysis on the first data set, and the D-items and G-items were specified as two clusters, respectively, in an analysis on the combined data set.

RESULTS

Preliminary data visualization

Except for item G6, normal quantile plots for the G-items indicate only a slight departure from normality in each item (Figure 9). Inter-relationships of the G-items are examined by a scatterplot matrix, as shown in Figure 10(a). In the pair, G1 and G6, at first glance, it seems that an outlier is present.

However, removing this observation does not have a substantive impact on the relationship between the two items. In Figure 10(b), which is a magnified image of the top, second from the right, in the scatterplot matrix, the black regression line results from using all observations whereas the red regression line is plotted without the suspected outliers. The two slopes are almost the same. Hence, the analysis proceeded without excluding any observations.

Figure 9. Distribution and normal quantile plots of G-items

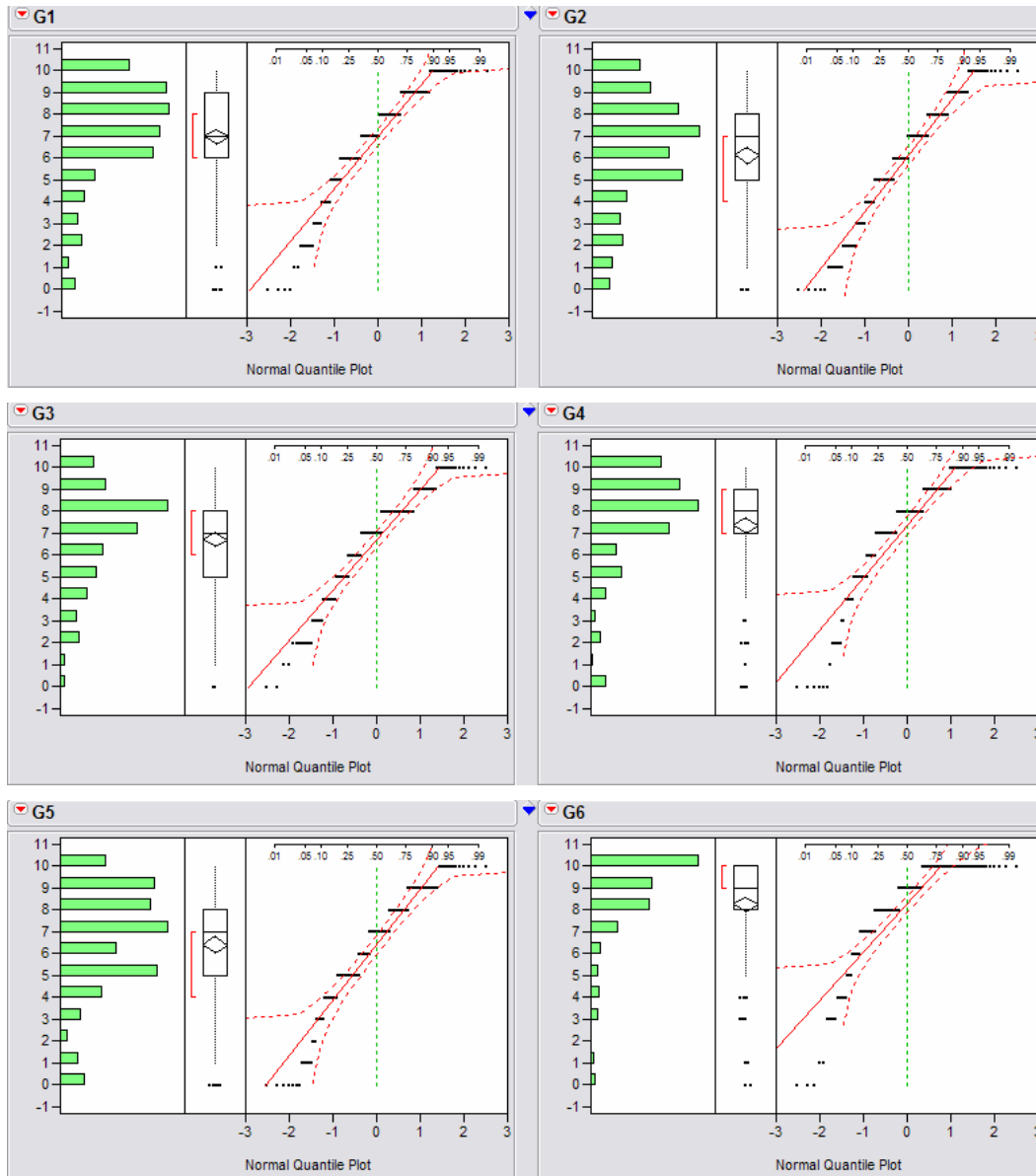


Figure 9 (continued). Distribution and normal quantile plots of G-items

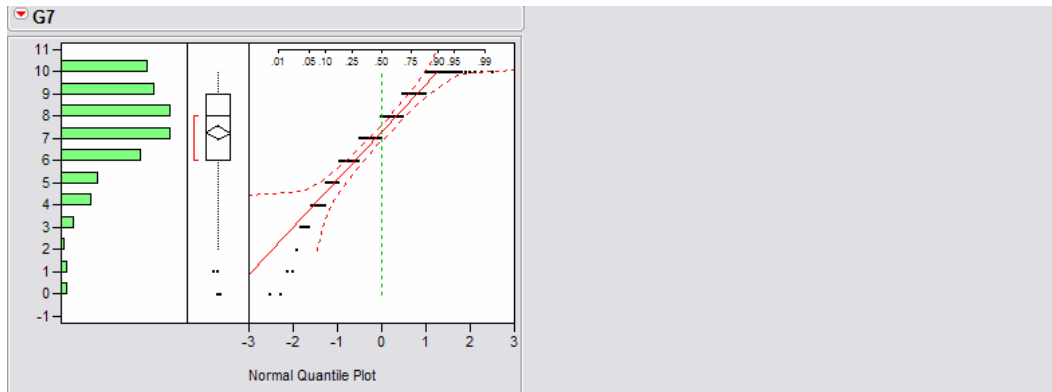


Figure 10(a). Scatterplot matrix of G-items.

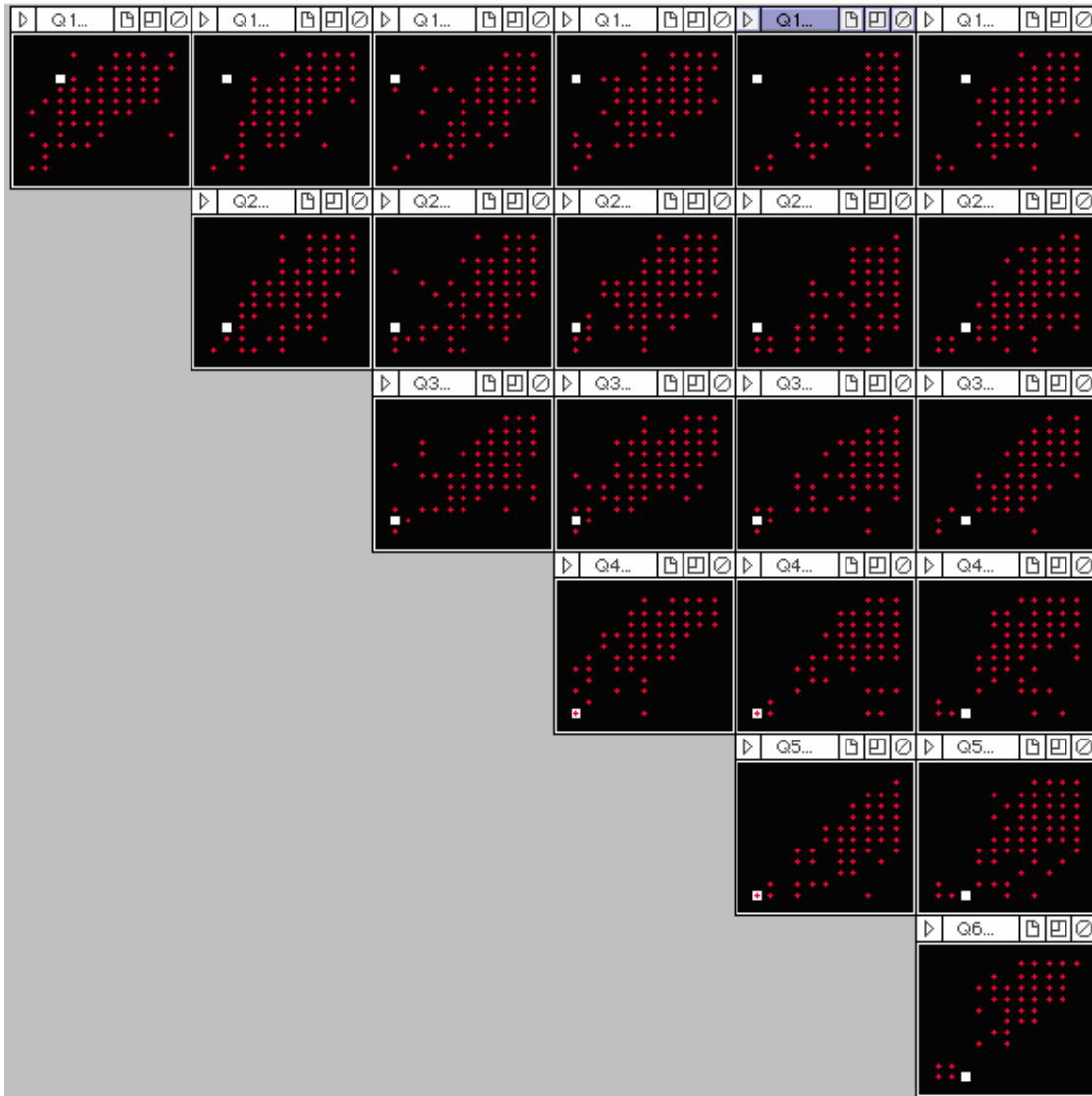
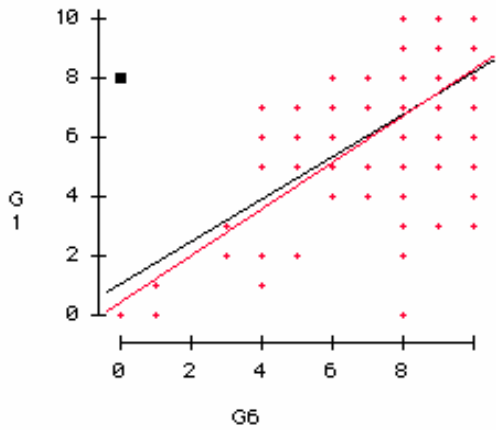


Figure 10(b) Scatterplot of G1 X G6



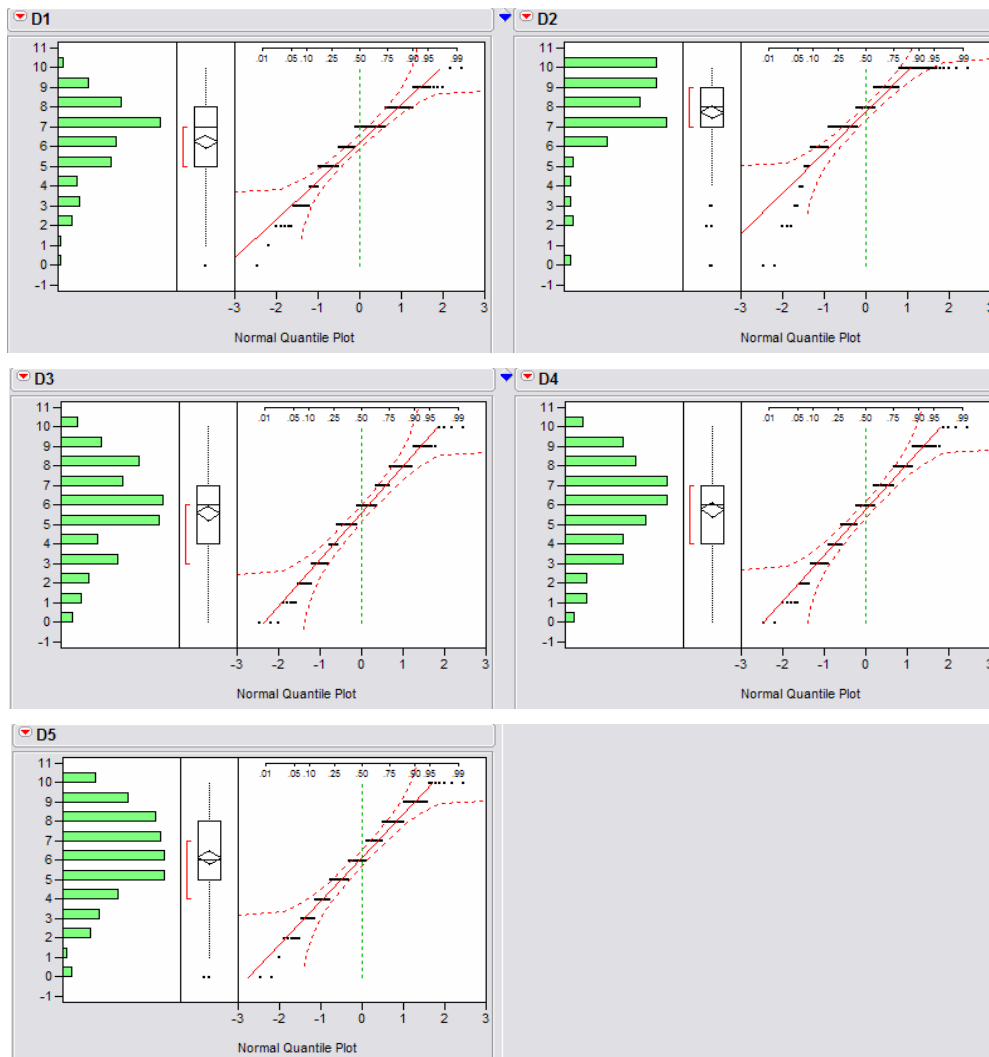
The data set that contains two underlying factors was also examined (see Figure 11). Except for D2, all item responses do not substantively depart from normality. Again, the scatterplot

matrix shows no outliers that would affect the pattern of relationships (Figure omitted, too dense to show).

Rasch analyses

Response probability threshold estimates were examined for the Rasch analyses conducted on each data set with RUMM. Most items displayed ordered thresholds. Item G7 showed a very small degree of threshold disorder, at the very lowest scale points, in both the G-items analysis and the combined data set analysis. D2 also displayed a small amount of threshold disorder in the combined analysis, again at the very lowest scale points. Threshold disorder at the very lowest scale point categories was not remarkable, given that both items also had very low response frequencies at the lowest scale points. A slight tendency toward a negative skew in the posttest responses for most items was somewhat more pronounced in these two items. This very low occurrence of threshold disorder did not lead us to consider the removal of any items from the analyses.

Figure 11. Distribution and normal quantile plots of D-items



For the single-factor data set, it was not surprising that the item residuals based on the gap between the expected and the observed in Chi-square tests yielded from RUMM did not raise any red flags (see Table 1). Please keep in mind that the Chi-square value generated for each item should not be used as a fitness index, since it is highly affected by the sample size. According to the conventional cut-off, a residual above 2 is considered problematic. Based on the visual inspection conducted in the dot plot (see Figure 12), an item with a residual substantively departing from the rest is treated as a misfit, but there are none in this result. The color codes indicate the values ranging from high to low based upon the color spectrum (near red=high, near cyan=low). Thus, inspection of residuals as well as the pattern of residuals shows

that all seven items contribute to the unidimensionality of the G-items.

When the combined data set, in which there are two correlated factors, was run in RUMM, examining the residuals by using a cutoff or a dot plot also does not issue any warning of the presence of any misfits and the violation of unidimensionality. However, when one looks closely at the sign of the residuals, all D-items have positive residuals whereas almost all G-items have negative residuals, except G5, which is close to zero. When the items are color-coded, the dot plot clearly reveals that there are two clusters (see Figure 13). One can indirectly infer from the residuals to the tacit implication that there may be two correlated yet distinct factors.

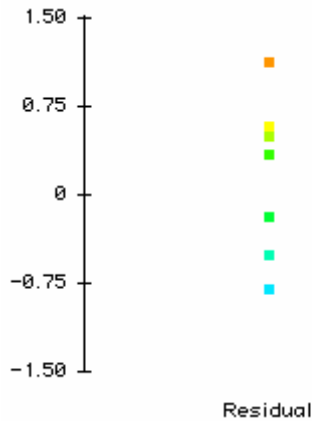
Table 1. Item parameters and residuals of G-items in RUMM

Item	Location	SE	Residual	Chi-Square	Probability
G1	0.090	0.07	-0.497	1.491	0.460
G2	0.615	0.06	-0.186	0.166	0.918
G3	0.113	0.07	0.510	0.213	0.896
G4	-0.125	0.07	0.351	1.233	0.528
G5	0.394	0.06	1.133	3.995	0.113
G6	-0.781	0.08	-0.800	0.123	0.939
G7	-0.305	0.07	0.587	0.158	0.922

Table 2. Item parameters and residuals of D-items and G-items in RUMM,

Item	Location	SE	Residual	Chi-Square	Probability
D1	0.233	0.07	1.685	2.178	0.319
D2	-0.357	0.07	1.310	3.273	0.173
D3	0.390	0.06	1.958	6.190	0.020
D4	0.345	0.06	0.169	1.117	0.561
D5	0.097	0.06	0.834	1.157	0.549
G1	-0.031	0.06	-1.129	1.002	0.595
G2	0.372	0.06	-0.886	0.279	0.866
G3	0.002	0.06	-0.335	0.785	0.667
G4	-0.176	0.06	-0.899	3.219	0.179
G5	0.223	0.06	0.072	0.135	0.933
G6	-0.748	0.07	-1.025	1.165	0.547
G7	-0.351	0.06	-0.930	0.986	0.600

Figure 12. Dot plot of residuals of G-items



alone. When the observed eigenvalue is much bigger than that arising from the chance hypothesis, it implies that the observation must arise from structure rather than by chance. However, please be cautioned that one should not blindly follow a criterion, no matter whether it is Kaiser criterion or PA criterion. When one looks carefully at the parallel scree plot, one can see that the actual eigenvalue of the seven-factor solution is also slightly higher than its resample counterpart. Needless to say, it would be absurd to adopt a seven-factor solution! Figure 15 (a), (b) and Table 5 show the result of parallel analysis using the dataset with two constructs. As expected, two factors are retained.

Parallel analyses

Figure 14(a) and (b) reveal the result of parallel analysis using the single construct, G-items dataset. In this parallel scree plot (a), the red line denotes the actual eigenvalues, the green line represents estimated eigenvalues at 95 percentile, and the gray line depicts the mean eigenvalues resulted from repeated sampling. This information is also reported in Table 4 whereas Figure 14(b) shows all eigenvalues yielded from re-sampling. The number of factors extracted should have eigenvalues greater than those generated from random matrix. This occurs in the first factor. It is worth noting that the actual eigenvalue of the first factor is far larger than the re-sampled one. This is due to the fact that the purpose of the re-sampling is to simulate a reference set by chance

Figure 13. Dot plot of residuals of G-items and D-items

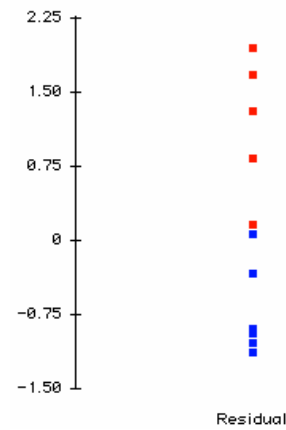


Figure 14(a) Parallel scree plot of G-items. (b) Boxplots of resampled eigenvalues

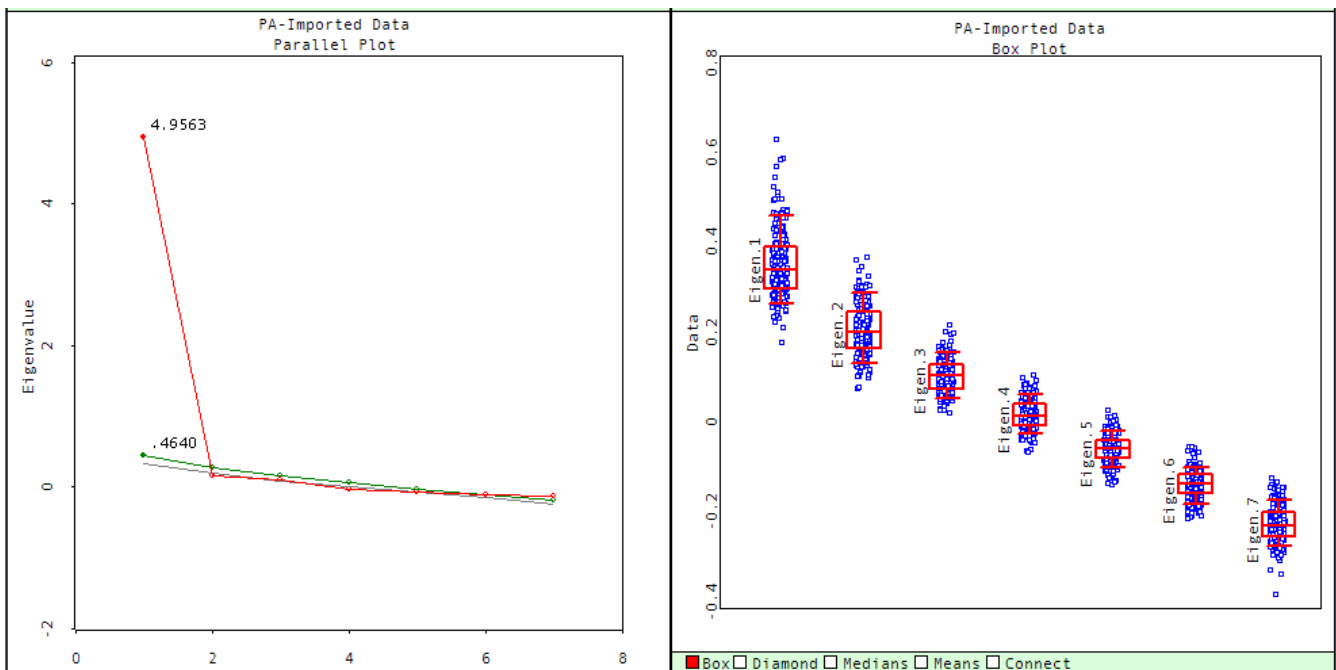


Table 4. Observed, mean resampled, and estimated eigenvalues of G-items

	Observed	Mean	95 percentile
Eigenvalue1	4.95635	0.35188	0.46399
Eigenvalue2	0.17779	0.20636	0.29323
Eigenvalue3	0.11011	0.10444	0.17517
Eigenvalue4	-0.02200	0.01784	0.06881
Eigenvalue5	-0.06410	-0.05923	-0.01089
Eigenvalue6	-0.10198	-0.13832	-0.08947
Eigenvalue7	-0.12031	-0.22763	-0.16596

Figure 15(a) Parallel scree plot of G-items and D-items. (b) Boxplots of resampled eigenvalues

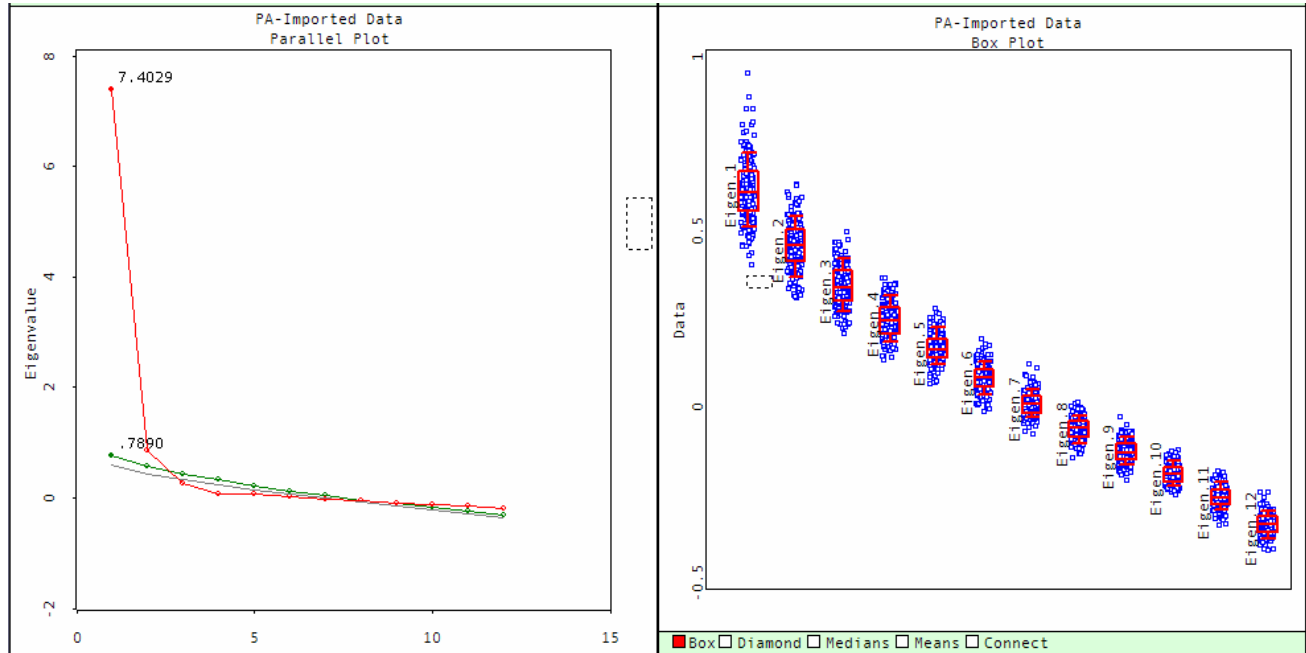


Table 5. Observed, mean resampled, and estimated eigenvalues of G-items and D-items

	Observed	Mean	95 percentile
Eigenvalue1	7.40291	0.61997	0.78897
Eigenvalue2	0.87641	0.46276	0.60354
Eigenvalue3	0.29306	0.34673	0.45366
Eigenvalue4	0.10394	0.25032	0.35727
Eigenvalue5	0.09489	0.16835	0.23571
Eigenvalue6	0.05389	0.08542	0.15040
Eigenvalue7	0.00701	0.01073	0.07231
Eigenvalue8	-0.01901	-0.06065	-0.01721
Eigenvalue9	-0.06914	-0.12550	-0.06900

Table 5 (continued). Observed, mean resampled, and estimated eigenvalues of G-items and D-items

	Observed	Mean	95 percentile
Eigenvalue10	-0.08933	-0.18860	-0.14759
Eigenvalue11	-0.12820	-0.25506	-0.21315
Eigenvalue12	-0.16091	-0.33225	-0.28272

TETRAD analyses

The result of TETRAD analysis using the data set with a one-dimensional factor structure is shown in Figure 16. The purify algorithm easily included all items in a single factor. Figure 17 shows the analysis result using the data set with two underlying factors. For the sake of experimentation, all items are forced into a single cluster. As a result, TETRAD selected items from the two underlying factors to formulate a single-dimensional factor model.

Figure 16. TETRAD graph of a single-factor model

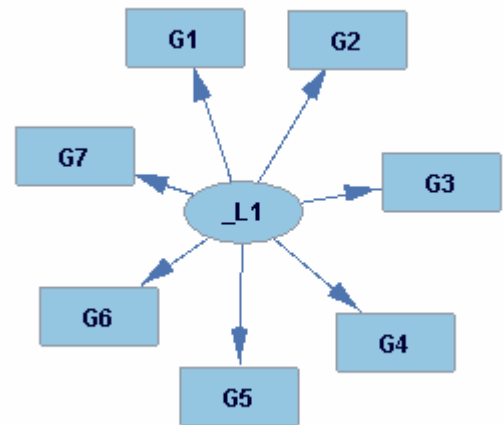
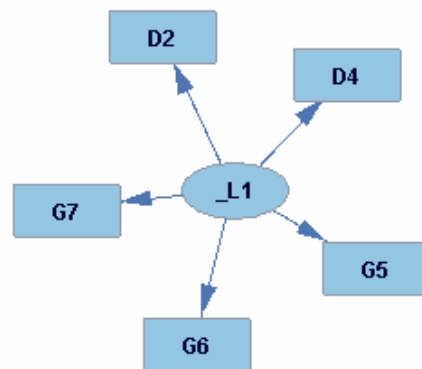
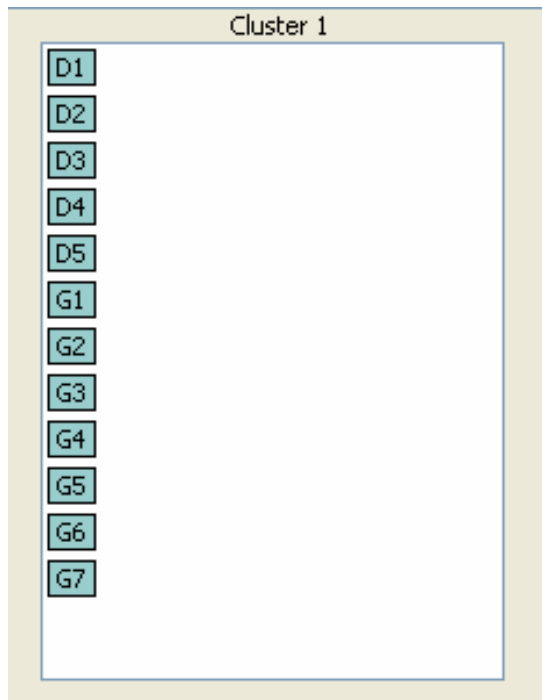


Figure 17. TETRAD graph of a single-factor model with D-items and G-items

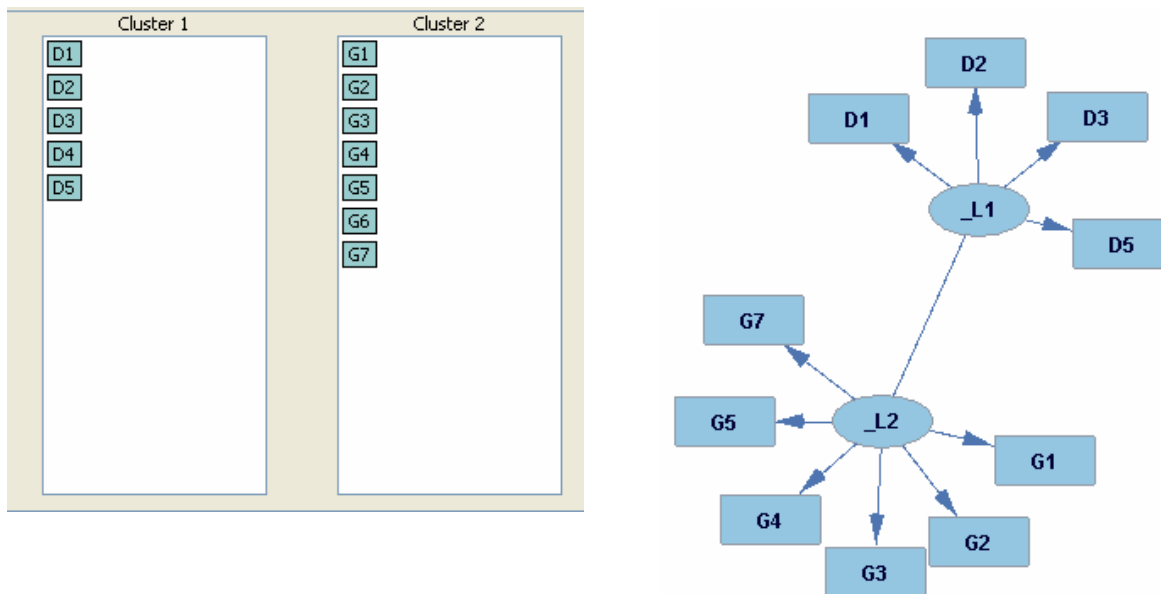


Please keep in mind that unlike EFA, TETRAD can act like CFA that allows the analyst to specify which items are loaded into which factors. When items were grouped into two clusters prior to applying the purify algorithm, TETRAD excluded D4 from the first factor and G6 from the second factor (see Figure 18). If one re-visits the histogram, boxplot, and the normal quantile plot shown in Figure 11, one can see that the distribution of D6 is the least normal. Nonetheless, why G4 is not included into the second factor requires further investigation. What the item author should do with G4 is debatable. It is important to point out that not only is CMU's TETRAD an exploratory technique, but also it is aligned to the reasoning of data mining (Yu, 2007). Both EDA and data mining maintains that strong assumptions based upon prior theory are discouraged. In other words, CMU's TETRAD endorses the role in theory in research design in the sense of determining which relevant variables should be included into the search space. But after the search algorithms are employed, the researcher should let the data speak for itself. In this case, the item author should drop G4. This decision is sound if TETRAD is the only employed method for this data set. However, this result is not corroborated by two other methods, which are also exploratory in essence.

DISCUSSION

It was found that preserving a unidimensional scale by removing misfits in Rasch modeling, extracting factors by comparing actual and re-sampled eigenvalues in PA, and purifying a single-factor model by purging impure indicators in TETRAD do not necessarily lead to the same conclusion, when blindly applying conventional assessment criteria. While the Rasch approach did not show the item misfit that would indicate deviation from clear unidimensionality, one cannot dismiss the usefulness of the Rasch approach in terms of detecting underlying dimensions. Even if no misfits are identified in the data set, the residual information provided hints to the researcher that different constructs may be entangled within the same survey or exam. PA seems to be a robust method for its success in identifying the correct number of factors. But the problem of dimensionality does not go away if the modeler simply switches from one set of criterion (eigenvalue greater than one) to another (actual eigenvalue greater than the resampled eigenvalue). One may obtain an absurd result if the PA criterion is blindly followed. TETRAD successfully confirmed one dimension in the single-construct data set and was able to confirm two

Figure 18. TETRAD graph of a two-factor model.



dimensions in the combined data set, yet excluded one item from each cluster, for no obvious reasons. The outcomes of all three approaches substantiate the conviction that the assessment of dimensionality requires a good deal of judgment.

Ambiguity may exist with respect to dimensionality. Despite the presence of two distinct factors within the combined data set, the moderately high degree of correlation between the two dimensions allowed for the property of unidimensionality to be supported within the Rasch

framework. If the interpretation and practical application of the scales involved allows for one broad dimension, defined, say, as a more general perception of conceptual understanding of mathematical concepts, then results that do not contradict unidimensionality are reasonable. In his comprehensive review of the methods and dimensionality indices available at the time, Hattie (1985) concluded that assessing unidimensionality may indeed require an act of judgment, and “even when there is an index, then judgment must still be used when interpreting it” (p. 159). Having a theoretical basis for test structure and multiple sources of evidence to support construct validity of the instrument is critical. The assessment of dimensionality should reflect instrument usage. If a single score is intended as the basis of inferences for each respondent, then unidimensionality must be established. If scores derived from multiple scales are to be used, then evidence to support multiple dimensions must be obtained.

Although the idea of TETRAD has been around since the turn of the last century, its implementation in computation is still fairly new. These examples show that while TETRAD was successful in confirming a single-factor model and implying a multi-dimensional model when a single-factor model is incorrectly specified, it excluded some items in the two-factor model without an obvious reason. Nevertheless, TETRAD is built upon philosophically rich assumptions, such as CCP, CMC, and FA. Also, TETRAD is a very user-friendly tool to conduct a dimensionality test in a CFA fashion. Researchers are strongly encouraged to explore its theoretical and practical potentials.

References

- Assessment Systems Corporation. (2007). *DIMPACK: Nonparametric dimensionality analysis package* [Computer software and Manual]. St Paul, MN: The Author.
- Andrich, D. (1978). A rating formulation for ordered categories. *Psychometrika*, *43*, 357-374.
- Andrich, D., de Jong, J. H. A. L. & Sheridan, B. E. (1997). Diagnostic opportunities with the Rasch model for ordered response categories. In J. Rost & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences* (pp. 59-70). Münster: Waxmann.
- Andrich, D., Lyne, A., Sheridan, B., & Luo, G. (1997) Rasch Unidimensional Measurement Models computer program. [Computer software and manual]. Australia: RUMM Laboratory.
- Bentler, P. M., & Wu, E. J.-C. (1993). *EQS/Windows user's guide*. Los Angeles: BMDP Statistical Software.
- Bock, D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement*, *12*, 261-280.
- Bollen, K. (1980). Issues in the comparative measurement of political democracy. *American Sociological review*, *45*, 370-390.
- Bollen, K., & Ting, K. (1993). Confirmatory tetrad analysis. *Sociological Methodology*, *23*, 147-175.
- Bollen, K., & Ting, K. (1998). Bootstrapping a test statistic for vanishing tetrads. *Sociological Methods and Research*, *27*, 77-102.
- Bollen, K., & Ting, K. (2000). A tetrad test for causal indicators. *Psychological Methods*, *5*, 3-22.
- Browne, M. W. (1984). Asymptotic distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62-83.
- Budescu, D. V., Cohen, Y., & Ben-Simon, A. (1997). A revised modified parallel analysis for the construction of unidimensional item pools. *Applied Psychological Measurement*, *21*, 233-252.
- Carroll, J. B. (1945). The effect of difficulty and chance success on correlation between items or between tests. *Psychometrika*, *10*, 1-19.
- DataDescription, Inc. (2007). DataDesk [Computer software and manual]. Ithaca, NY: The author.
- Druzdzel, M, & Glymour, C. (1995, March-April). *Having the right tool: Causal graphs in teaching research design*. Paper presented at the University of Pittsburgh Teaching Excellence Conference, Pittsburgh, PA.
- Gibbons, R. D.; Bock, D., Hedeker, D., Weiss, D., Segawa, E., Bhaumik, D. K., Kupfer, D. J., Frank, E, Grochocinski, V. J., Stover, A. (2007). Full-information item bifactor analysis of graded response data. *Applied Psychological Measurement*, *31*, 4-19.
- Gierl, M. J., Leighton, J. P., & Tan, X. (2006). Evaluating DETECT classification accuracy and consistency when data display complex structure. *Journal of Educational Measurement*, *43*, 265-289.
- Glymour, C. (1982). Casual inference and causal explanation. In R. McLaughlin (Ed.), *What? Where? When? Why? Essays on induction, space, and time, explanation* (pp. 179-191). Boston, MA: D. Reidel Publishing Company.
- Glymour, C. (2001). *Mind's arrows: Bayes and graphical causal models in psychology*. Cambridge, MA: MIT Press.
- Glymour, C., Scheines, R., Spirtes, P., & Kelly, K. (2000). *Discovering causal structure*. Orlando, FL: Academic Press.
- Hart, B., & Spearman, C. (1913). General ability, its existence and nature. *British Journal of Psychology*, *5*, 51-84.

- Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement, 9*, p. 139-164.
- Hayton, J. C., Allen, D. G. & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7*, 191-205.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179-185.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24*, 417-441, 498-520.
- Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications*, 2nd ed. Chicago: SPSS, Inc.
- Kim, H.R. (1994). *New techniques for the dimensionality assessment of standardized test data*. Unpublished doctoral dissertation, University of Illinois at Urbana—Champaign.
- Kline, R. (2006). Reverse arrow dynamics: Formative measurement and feedback loops. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 43-68). Greenwich, CT: Information Age Publishing.
- Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method*. London: Butterworth.
- Ledesma, D. & Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out Parallel Analysis. *Practical Assessment, Research and Evaluation, 12*. Retrieved May 4, 2007 from <http://pareonline.net/pdf/v12n2.pdf>.
- Li, Q., & Reuveny, R. (2003). Economic Globalization and Democracy: An Empirical Analysis. *British Journal for Political Science, 33*, 29-54.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. New York: Erlbaum Associates.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test errors*. Reading, MA: Addison-Wesley.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.
- Masters, G. N. & Wright, B. D. (1997). The partial credit model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 101-121). New York: Springer.
- McDonald, R. P. (1967). *Nonlinear factor analysis* (Psychometric Monographs, No. 15). The Psychometric Society.
- McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology, 34*, 100-117.
- McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Lawrence Erlbaum.
- McDonald, R. P. (1997). Normal-ogive multidimensional model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory*. New York: Springer.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.
- Meara, K., Robin, F., & Sireci, S. (2000). Using multidimensional scaling to assess the dimensionality of dichotomous item data. *Multivariate Behavioral Research, 35*, 229-259
- Meiser, T., Stern, E., & Langheine, R. (1998). Latent change in discrete data: Unidimensional, multidimensional, and mixture distribution Rasch models for the analysis of repeated observations. *Methods of Psychological Research, 3*. Retrieved August 27, 2007 from <http://www.mpr-online.de/issue5/art6/article.html>.
- Mulaik, S., & Millsap, R. (2000). Doing the Four-Step Right. *Structural Equation Modeling, 7*, 36-73.
- Muthén, L. K., & Muthén, B. (1998). *Mplus: The comprehensive modeling program for applied researchers: User's guide*. Los Angeles: Muthén & Muthén.
- Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research, 1960. Expanded edition, Chicago: The University of Chicago Press, 1980.
- Reichenbach, H. (1956). *The direction of time*. Berkeley, CA: University of Los Angeles Press.
- Roznowski, M., Tucker, L. R., & Humphreys, L. G. (1991). Three approaches to determining the dimensionality of binary items. *Applied Psychological Measurement, 15*, 109-127.
- Roussos, L.A., Stout, W.F., & Marden, J.I. (1998). Using new proximity measures with hierarchical cluster analysis to detect multidimensionality. *Journal of Educational Measurement, 35*, 1-30
- SAS Institute. (2007). JMP [Computer software and manual]. Cary, NC: The author.
- Scheines, R., Glymour, C., & Spirtes, P. (2005). The TETRAD project: Causal models and statistical data [Computer software and manual]. Retrieved October 13, 2005, from <http://www.phil.cmu.edu/projects/tetrad/>

- Scheines, R., Spirtes, P., Glymour, C., Meek, C., & Richardson, T. (1998). The TETRAD project: Constraint based aids to causal model specification. *Multivariate Behavioral Research, 33*, 65-117.
- Shipley, B. (2000). *Cause and correlation in biology: A user's guide to path analysis, structural equations and causal inference*. Cambridge: Cambridge University Press.
- Smith, R. M. (1996). A comparison of methods for determining dimensionality in Rasch measurement. *Structural Equation Modeling, 3*, 25-40.
- Smith R. M., Miao C.Y. (1994). Assessing unidimensionality for Rasch measurement. In M. Wilson (Ed.), *Objective measurement: Theory into practice (Vol. 2)* (pp. 316-327). Norwood, NJ: Ablex.
- Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, Prediction and Search*. Cambridge: MIT Press.
- Stiglitz, J. E. (2002). *Globalization and its discontents*. New York: W.W. Norton & Co..
- Stout, W. F. (1987). A nonparametric approach for assessing latent trait dimensionality. *Psychometrika, 52*, 589-617.
- Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensional assessment and ability estimation. *Psychometrika, 55*, 293-326.
- Stout, W. F., Nandakumar, R., & Habing, B. (1996). Analysis of latent dimensionality of dichotomously and polytomously scored test data. *Behaviormetrika, 23*, 37-65.
- Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika, 52*, 393-408.
- Tate, R. L. (2003). A comparison of selected empirical methods for assessing the structure of responses to test items. *Applied Psychological Measurement, 27*, 159-203.
- Tennant, A., & Pallant, J. (2006). Unidimensionality matters. *Rasch Measurement Transactions, 20*, 1048-1051.
- Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 41-71). Boston: Kluwer Academic.
- Waugh, R. F., & Chapman, E. S. (2005). An analysis of dimensionality using factor analysis (true-score theory) and Rasch measurement: What is the difference? Which method is better? *Journal of Applied Measurement, 6*, 80-99.
- Weng, L.J., & Cheng, C. P. (2005). Parallel analysis with unidimensional binary data. *Educational and Psychological Measurement, 65*, 697-716.
- Wright, B. D. (1997). Rasch factor analysis. In M. Wilson, G. Engelhard, & K. Draney (Eds.), *Objective measurement: Theory into practice (Vol. 4)* (pp. 113-137). Norwood, NJ: Ablex.
- Wright, B. D., & Panchapakesan, N. (1969). A procedure for sample-free item analysis. *Educational and Psychological Measurement, 29*, 23-48.
- Young, F. (1999). VisTa: The visual statistics system [Computer software and manual] Retrieved March 2006, from <http://www.uv.es/prodat/ViSta/>
- Yu, C. H. (2007). Is automated data mining an emerging paradigm that can supersede hypothesis testing? *Theory and Science, 9*. Retrieved August 30, 2007 from <http://theoryandscience.icaap.org/content/vol9.2/Chong.html>.
- Zhang, J., & Stout, W.F. (1999). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika, 64*, 213-249
- Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological bulletin, 99*, 432-442.

Note

¹ Although DIMTEST has been widely regarded as a versatile tool that performs well in most situations, previous studies found that DIMTEST has some trouble in detecting multidimensionality when the number of test items is small (Meara, Robin & Sireci, 2000) while some studies cannot verify whether DIMTEST is superior to conventional approaches (Tate, 2003). DETECT is most useful when the data display approximate simple structure. However, when the underlying structure is complex and the correlation between dimensions is very high, DETECT does not work well with any complex dimensional structure (Gierl, Leighton, & Tan, 2006). To compensate for the preceding shortcomings, HCA/CCPROX can be used to conduct a latent multidimensionality-sensitive hierarchical cluster analysis on dichotomously scored items. Recently the above three software modules are bundled as DIMPACT with many enhancements (Assessment Systems Corporation, 2007), and the efficacy of this complementary software modules await evaluation.

Andrich, de Jong, and Sheridan (1997) and Masters and Wright (1997) have presented differing views regarding the interpretation of threshold parameters. Andrich, de Jong, and Sheridan (1997) regard the increasing value of threshold parameters as an implication of hypothesized ordered categories. Masters and Wright (1997) contend that the parameters can take any order, as in an achievement item that requires multiple steps, but not necessarily in order. Meiser, Stern, and Langeheine (1998) present a mediating view that recommends interpretation consistent with the nature of the response categories.

Citation

Yu, Chong Ho, Sharon Osborn Popp, Samuel DiGangi, and Angel Jannasch-Pennell (2007). Assessing unidimensionality: A comparison of Rasch Modeling, Parallel Analysis, and TETRAD. *Practical Assessment Research & Evaluation, 12*(14). Available online: <http://pareonline.net/getvn.asp?v=12&n=14>

Correspondence

Chong Ho Yu, Ph.Ds
PO Box 612
Tempe AZ 85280
Chonghoyu [at] gmail.com