Comparison of Various Polytomous Item Response Theory Modeling Approaches for Task-Based Simulation CPA Exam Data

AICPA 2014 Summer Internship Project

Oksana Naumenko

The University of North Carolina at Greensboro

Table of Contents

Introduction

Unidimensional Item Response Theory (IRT) models are frequently used for calibration of item responses in educational assessment. Two of the necessary and related assumptions imposed by IRT are unidimensonality and local item independence, the notions that a given assessment is measuring one, and only one dominant construct and that items are not related above and beyond this target construct. When tests are composed of small subtests that are large enough to carry their own context (Wainer & Kiely, 1987; Wainer & Lewis, 1990), (i.e., testlets), scores tend to indicate that local item dependence, and consequently, unidimensionality within testlets, are likely to be violated (Lee, Kolen, Frisbie, & Ankenmann, 2001; Wainer & Thissen, 1996). Using dichotomous unidimensional IRT (DIRT) models, which inherently assume such independence, to score and equate testlet-based tests may therefore be problematic (Lee et al., 2001). Previously, testlet scores obtained from polytomous unidimensional IRT (PIRT) models have been found more appropriate than DIRT models for eliminating the influence of dependence between items within a testlet (Lee et al., 2001; Sireci, Thissen, & Wainer, 1991).

Treating testlet items as independent is problematic in calculation of reliability (Sireci, Thissen, & Wainer, 1991; Thissen, Steinberg, & Mooney, 1989; Wainer, 1995; Wainer & Thissen, 1996). Items exhibiting local dependence are essentially redundant, as they provide very similar information, and thus do not provide additional information about an examinee's ability (Wainer, 1995; Wainer & Thissen, 1996). This finding is especially deleterious to the ultimate inferences drawn from high-stakes assessment scores, which may be used to classify examinees according to a pass/fail cutoff. As one example, Wainer (1995) used Bock's (1972) nominal response PIRT model to show that the testlet-based reliability of two LSAT sections was considerably lower than the traditional calculation of

reliability due to unmodeled local item dependencies. Additionally, differential item functioning (DIF) was shown to be exposed on the testlet level, and not the individual item level. In an attempt to alleviate similar issues, Thissen and colleagues (1989) compared Bock's (1972) model to the 3PL DIRT model, and also found that the latter overestimated the precision of measurement. Sireci, Thissen, and Wainer (1991) pointed out that when local independence only holds between testlets, then appropriate calculations of reliability should be based on testlets as the unit of measurement.

The overarching goal for the current Summer Internship project is to compare the fit of PIRT models for task-based CPA Exam data. As part of the project, models for polytomous item responses and goodness of fit methods for the Graded Response Model (GRM) and the Generalized Partial Credit Model (GPCM) were reviewed. The review serves the purpose of informing the subsequent analyses that will compare (i) theta estimates under each model to theta estimates under the 3PL model, and (ii) the theoretical and statistical fit between models, and (iii) the shape of test information functions between models. The following literature review includes (i) a broad overview of models available for polytomous item responses, (ii) the background of the GRM and GPCM, and (iii) the methodology for assessing goodness of fit to the data when modeling with the GRM and GPCM.

Review of Models for Polytomous Item Responses

Currently, the CPA exam task-based simulations (TBSs) are scored using the 3PL IRT model using operational item parameters. Specifically, between one and thirteen measurement opportunities (MOs) are dichotomously scored within each TBS. The 3PL is characterized by three item parameters, reflecting item difficulty, item discrimination, and pseudo-guessing. The model is specified as follows:

$$P(X_{ie}=1|\,\theta_e,\beta_i,\alpha_i,\gamma_i)=\gamma_i+(1-\gamma_i)\frac{\exp[\alpha_i(\theta_e-\beta_i)]}{1+\exp[\alpha_i(\theta_e-\beta_i)]},$$

where X_{ie} is the response of candidate *e* to item *i*, θ_e is the trait level for candidate *e*, β_i is difficulty of item *i*, α_i is discrimination for item *i*, and γ_i is the lower asymptote (pseudo-guessing) for item *i*. With varying scoring rubrics applied to the scoring of each MO in a TBS, the probability of being able to guess the correct answer likely varies with each dichotomously scored MO. Thus, modeling the guessing parameter is important given that TBS information contributes to the candidate's final score.

A scoring method that is consistent with PIRT models is summing across the *n* MOs available within each TBS. The resulting score is a number correct for a given TBS; however, it should be noted that the response pattern contains further information and may also be considered and scored using a multicategorical PIRT model. Since the development of the first models for polytomously-scored candidate responses (e.g., Bock, 1972; Samejima, 1969; 1972), numerous PIRT models have been introduced and popularized in certification practice. The following review assumes the basic understanding of the dichotomous IRT models, and presents an account of several widely-used PIRT models.

In the context of certification exams, polytomous items are cognitive and non-cognitive stimuli that have the potential of having more than two score outcomes. IRT applied to polytomous items specifies an item response function (IRF) for each possible outcome. An IRF specifies the probability of an outcome Y_i as a function of the target trait. Unique to PIRT models is the step function (Masters, 1982), or transitional models that specify a wide range of IRFs using some number of item parameters. Various PIRT models can be specified based on how step functions are defined and used to interpret the probability

of a response category. Using this structure, major polytomous response models are described next.

The first group of models can be described as using adjacent categories only in calculating the probability of failure or success on the item. Specifically, these models define the k^{th} step function using only the adjacent score categories (e.g., $Y_i = 0$ and $Y_i = 1$), and specify the probability of success on the first step as the probability that $Y_i = 1$ given that $Y_i = 0$ or $Y_i = 1$. The subsequent steps are interpreted similarly. Models under the "divide-by-total" approach (Thissen & Steinberg, 1986) are the Partial Credit Model (PCM; Masters, 1982), the Generalized Partial Credit Model (GPCM; Muraki, 1992; 1993), and the Rating Scale Model (RSM; Andrich, 1978a; 1978b).

Partial Credit Model

The PCM uses the Rasch model to specify the probability of success at k^{th} step such that the IRF for $Y_i = 0$ has the form

$$P_{i0}(\theta) = \frac{1}{1 + \sum_{r=1}^{m} [\exp \sum_{k=1}^{r} (\theta - b_{ik})]}$$

and the IRF for $Y_i = j > 0$ have the form

$$P_{ij}(\theta) = \frac{\exp[\sum_{k=1}^{j} (\theta - b_{ik})]}{1 + \sum_{r=1}^{m} [\exp \sum_{k=1}^{r} (\theta - b_{ik})]}$$

where step is denoted by $r = 1, 2, 3, m$ and j represents the score category. Thus, for a set of n items there will be n x m item parameters.

The PCM is quite popular in assessment contexts due to its parsimonious nature. Because the PCM allows for a relatively small number of estimates per set of items, sample sizes as small as 300 return stable item parameter and trait estimation (de Ayala, 2009).

Generalized Partial Credit Model

Unlike the PCM, the GPCM includes the item-level discrimination parameter and expresses the IRF for $Y_i = 0$ as

$$P_{i0}(\theta) = \frac{1}{1 + \sum_{r=1}^{m} (\exp \sum_{k=1}^{r} [a_i(\theta - b_{ik})])}$$

and the IRF for $Y_i = j > 0$ have the form

$$P_{ij}(\theta) = \frac{\exp(\sum_{k=1}^{j} [a_i(\theta - b_{ik})])}{1 + \sum_{r=1}^{m} (\exp \sum_{k=1}^{r} [a_i(\theta - b_{ik})])}$$

where a_i is the item discrimination parameter common across all m steps, but unique to each item. For a set of n items, $n(m+1)$ parameters are estimated.

The GPCM is the most general of the three "divide-by-total" PIRT models; fixing the value of a_i to 1 across items reduces to the PCM. The GPCM is flexible in that it allows the possibility of identifying item response options that may be redundant with each other. For example, IRFs for some response options may be centered at the same ability estimate.

Rating Scale Model

One difference between the PCM and RSM is that the RSM constrains the distance between the item difficulty values to be the same for all items on the instrument, such as when item responses are elicited by a common set of behavioral anchors (i.e., Likert-type anchors). Another difference here is the parameterization that uses *threshold* and *distance* parameters. A threshold parameter can be conceptualized as "average difficulty", or an item center around which IRFs form. Specifically, the value of d_i is the center value of a target trait so that the average distance between the m values of b_{ik} and d_i for an item is zero. Each distance from d_i is $t_{ik} = b_{ik} - d_i$. The RSM IRF for $Y_i = 0$ is

$$P_{i0}(\theta) = \frac{1}{1 + \sum_{r=1}^{m} [\exp \sum_{k=1}^{r} (\theta - d_i - t_k)]}$$

and the IRF for $Y_i = j > 0$ have the form

$$P_{ij}(\theta) = \frac{\exp[\sum_{k=1}^{j} (\theta - d_i - t_k)]}{1 + \sum_{r=1}^{m} [\exp \sum_{k=1}^{r} (\theta - d_i - t_k)]}$$

For a set of n items, the RSM only estimates $n+m$ parameters, thus allowing for smaller sample sizes to be used in estimation (de Ayala, 2009).

Another set of polytomous response model approaches uses different numbers of response categories depending on which step function is in question. Specifically, these models define the k^{th} step as advancing to score category k or higher ($Y_i \geq k$) given $Y_i \geq k - 1$. Therefore, success is defined as $Y_i \geq k$ and failure – as $Y_i = k - 1$. *Continuation Ratio* models are theoretically justified when the score categories are assumed to have an underlying sequential process, such as when the score categories reflect the successful application of a hierarchically ordered set of skills or processes (Penfield, 2014).

Sequential Rasch Model

The SRM substitutes the Rasch model for success across m steps. Specifically, the SRM-specified IRF for $Y_i = 0$ is

$$P_{i0}(\theta) = \frac{1}{1 + \exp(\theta - b_{i1})}$$

and the IRF for $y_i = j > 0$ is

$$P_{ij}(\theta) = \frac{\prod_{k=1}^{j} \exp(\theta - b_{ik})}{\prod_{k=1}^{j+1} [1 + \exp(\theta - b_{ik})]}$$

where $\exp(\theta - b_{ik}) = 0$ for $k = m + 1$.

The interpretation of the b_{ik} parameter is the value of theta at which the height of the IRF for $Y_i = k$ - 1 equals the sum of the heights of the IRFs for $Y_i \geq k$.

In the "difference models" (Thissen & Steinberg, 1986) IRT model category, or cumulative approach (Penfield, 2014) to defining step functions, the k^{th} step function describes failure as $Y_i < k$ and a success as $Y_i \geq k$. Thus, all score categories are used in quantifying the probability of success or failure. The Graded Response Model (GRM; Samejima, 1969) belongs to this cumulative approach.

Graded Response Model

The GRM approximates probabilities based on the 2PL specification such that separate b_{ik} parameters are estimated for each step of the item, and one a_i parameter is used for all steps for each item. The GRM specifies m - 1 "boundary" response functions that indicate the cumulative probability for a response category greater than the option of interest. The equation for such a BRF is closely related to the 2PL logistic model for dichotomous response data:

$$P_{ij}^{*}(\theta) = \frac{\exp[a_i(\theta - b_{ij})]}{1 + \exp[a_i(\theta - b_{ij})]}$$

However, the GRM is an "indirect" model in that the probability of responding to each category is captured by obtaining the IRFs from the difference between adjacent step functions. The b_{ik} are interpreted as the target trait value at which $P_{i0}(\theta) = .5$, b_{im} as the target trait value at which $P_{im}(\theta) = .5$, and for values in between steps $(b_{ik} + b_{ik} + 1)/2$ corresponds to the modal point of the IRF for $Y_i = k$ (Penfield, 2014). The justification for using GRM, or any model based on ordered response categories, with testlet-based scores is that testlet-based scores can theoretically have an ordered quality if they "correspond to the extent of completeness of the examinee's reasoning process within a testlet" (Lee et al.,

2001). That is, the more dichotomously-scored measurement opportunities within one testlet are answered correctly by an examinee, the more extensive is her ability.

Nominal Response Model

In the situation where item response options are not necessarily ordered in a pre-specified ways, nominal, or multiple-choice models are used to characterize item responses. The Nominal Response Model (NRM; Thissen & Steinberg, 1986; Bock, 1972) has been developed to describe the probability of a candidate responding in one of the available categories provided by an item (i.e., MCQ, or Likert-type item). The NRM has been applied most often in testlet applications with MCQ items (Sireci et al., 1991, Wainer, 1995). The idea of guessing in the case of the CPA Exam TBS is conceptualized somewhat differently because there is a varying probability of guessing per measurement opportunity. Nevertheless, in the NRM, the IRF for $Y_i = 0$ is defined as

$$P_{i0}(\theta) = \frac{1}{1 + \sum_{k=1}^{m} \exp(c_{ik} + a_{ik}\theta)}$$

And the IRF for $y_i = j > 0$ is

$$P_{ij}(\theta) = \frac{\exp(c_{ik} + a_{ik}\theta)}{1 + \sum_{k=1}^{m} \exp(c_{ik} + a_{ik}\theta)}$$

where c_{ik} is a location parameter such that the intersection of IRFs for $Y_i = 0$ and $Y_i = k$ is at theta = $-c_{ik}/a_{ik}$. Thus, for each item there are $2m$ item parameters. Several other versions of the NRM have been proposed to account for guessing behavior in candidates with low target ability (e.g., Revuelta, 2005; Thissen, Steinberg, & Fitzpatrick, 1989).

One of the goals of this project is to advise potential selection of a PIRT model that theoretically satisfies the assumptions underlying CPA Exam TBS item responses. Several

criteria have been outlined in the past (e.g., Ostini & Nering, 2005). Data characteristics and mathematical criteria are two criteria relevant to this study.

Considering data characteristics, because TBS data do not consistently conform to a multiple-choice format, the NRM, or other multiple-response models, are inappropriate. Moreover, data are not continuous, but ordered and have different numbers of categories, which theoretically precludes the use of a rating scale model. Remaining choices include adjacent category (i.e., GPCM) and cumulative boundaries (i.e., GRM) models. Samejima (1996) provided specific mathematical criteria justifying the fidelity between the psychological process of response production and the measurement model. Mathematical criteria involve several types of model fit measures, which have certain advantages and disadvantages outlined next.

Goodness of Fit Methods for GRM and GPCM

The two focal polytomous models of interest to the current project are the GRM (Samejima, 1969) and the GPCM (Muraki, 1992). The GRM and GPCM differ in the nature in which the IRFs are represented. The GRM manifests as a proportional odds model in which for each item, all response categories are collapsed into two categories when estimating the IRFs (Kang, Cohen, & Sung, 2009). As described above, a series of 2PL models are used in GRM item parameter calibration. On the other hand, for adjacent odds models like the GPCM, the focus is on the relative difficulty of each step needed to transition from one category to the next in an item score. Therefore, the two models do not indicate the same ordering among score categories and do not produce directly comparable parameters (Ostini & Nering, 2005), although many have found that these common polytomous IRT models tend to produce very similar results (Maydeu-Olivares, Drasgow, & Mead, 1994).

Some approaches to estimating the fit of a model are excluded from this review. Particularly, *residual-based measures* that evaluate the differences between observed and expected item responses and apply to only Rasch forms of PIRT models. The focus of the review will be on chi-square goodness-of-fit tests that are beyond residual-based tests.

Assessing item fit of an IRT model can be outlined in a few general steps when following the frequentist approach (Stone & Hansen, 2000). First, item and ability parameters are estimated under the chosen model. Then, candidates may be classified into several (e.g., 10) homogenous groups for which an observed score distribution is constructed by cross-classifying candidates using their ability estimates and score responses. A predicted score response distribution across score categories is constructed for each item. Discrepancies between the observed and predicted responses are then quantified and evaluated. Several item fit evaluation approaches exist that vary in the way candidates are grouped, the calculation of the expected values, and the determination of the chi-square statistic. In the following sections the nature of each available fit index for the GPCM and GRM is introduced, and studies behind the adequacy of each index are presented. Further, additional indices falling outside of the chi-square tradition are described.

Classical Goodness of Fit Tests

Bock's (1972) Pearson goodness of fit test. Bock's procedure involves subdividing the ability scale into *k* subgroups of similar sizes. The observed score distribution is then obtained by cross-classifying an individual's score response with the discrete ability scale. Predicted values for each ability interval consist of the median (or group centroid) of within-interval item parameter estimates. A chi-square statistic can then be calculated for each ability category *h* and response category *k*:

$$BCHI = \sum_{h=1}^{G} \sum_{k=0}^{m_i} \frac{N_h (O_{ihk} - E_{ihk})^2}{E_{ihk}(1 - E_{ihk})}$$

where N_h is the number of candidates with ability estimates falling within interval h, O_{ihk} is the observed proportion of candidates in interval h on item i with selected response category k, and E_{ihk} is the median proportion of candidates in interval h scoring in category k. Bock's (1972) procedure adjusts the degrees of freedom for uncertainty in estimated item parameters but not in ability estimates (Stone, 2000). BCHI was found to produce the fewest Type I errors of misfit when the generating model differed from the calibrating model (McKinley & Mills, 1985; Stone & Hansen, 2000), compared with Yen's (1981), McKinley & Mills's (1985) likelihood ratio, and a modified version of Wright and Mead's (1977) chi-square statistics.

Yen's (1981) Q_1 Chi-square statistic. Yen's (1981) Q_1 index is very similar to BCHI, with the exceptions that it specifies $h = 10$ ability intervals and uses the mean of the predicted probabilities within an interval. Like the BCHI, Q_1 also adjusts degrees of freedom (10) for the uncertainty in estimated item parameters (e.g., by subtracting 2 for the two-parameter logistic model) (DeMars, 2005; Stone, 2000). Stone and Hansen (2000) examined a Pearson's chi-square index similar to Yen's with real GRM-estimated data, and found extremely inflated Type I error rates, especially for short constructed-response tests (8, 16 items). Again, the flaw with this type of chi-square statistic is that examinees are grouped into intervals based on their IRT θ estimates, not their true θ, which inflates Type I errors (Orlando & Thissen, 2000).

***G^2* (McKinley & Mills, 1985).** A likelihood-ratio chi-square test was introduced by McKinley & Mills (1985), a version of which can be obtained through PARSCALE (Muraki & Bock, 1997). Here,

$$G_i^2 = 2\sum_{h=1}^{H_i}\sum_{k=0}^{m_j} r_{ihk} \times \ln \frac{r_{ihk}}{N_{ih} \times P_{ik}(\bar{\theta})}$$

where H_i is the number of ability intervals for item i, m_j is the number of response categories for item i, r_{ihk} is the number of observed candidates with a response category k in interval h, N_{ik} is total number of candidates in group k, and $P_{ik}(\theta)$ is the probability of response category k on item i, estimated by the item response function at the mean ability of candidates in interval h. Degrees of freedom are equal to the number of score intervals (H_i), but often are adjusted to $H_i - p$, where p is the number of estimated parameters per item. It is important to note that the PARSCALE index differs from the original G^2 in that no df adjustments are made for the uncertainty in either item or ability estimates (DeMars, 2005). DeMars generated normal and uniform ability distributions and fit the GRM and PCM to data from various test lengths and found that when the test length was 20, the Type I error rate for the PARSCALE index was stable for PCM (and close to nominal α, as would be expected) regardless of degrees of freedom. Type I error rates were inflated for both shorter and longer test lengths when the ability distribution was uniform for the GRM. α was inflated when one or more response categories were used infrequently. In the other conditions, the Type I error rate decreased as the degrees of freedom increased. Kang and Chen (2008) found similar results for the PARSCALE G^2.

Again, the issue of grouping examinees based on θ and disagreement regarding appropriate degrees of freedom lead to comparison of observed data to potentially inappropriate chi-square distributions.

Orlando and Thissen's (2000, 2003) *S*-X^2 generalization to polytomous models. Kang and Chen (2008) generalized Orlando and Thissen's (2000, 2003) *S*-X^2 statistic for use with polytomous response items as:

$$S - X^2 = \sum_{h=m_i}^{F-m_i} \sum_{k=0}^{m_i} N_h \frac{(O_{ihk} - E_{ihk})^2}{E_{ihk}}$$

where m_i is the highest response category for item i, k indicates the response category, F is the sum of m_i, h is a homogenous group of candidates, N_h is the number of candidates in group h, and O_{ihk} and E_{ihk} are the observed and predicted proportions of the k category response in item i for group h. The main advantage of S-X^2 is that, in contrast to *BCHI*, Q_1 and G^2, homogenous groups of candidates are based on observed test scores rather than model-based abilities. The reason for the summation for h from m_i through F-m_i is that within some groups with extremely low or high test scores, the E_{ihk} for some categories are always zero. To counteract this, such groups are collapsed with groups with $h = m_i$ or $h = F - m_i$. The expected category proportions E_{ihk} are computed using

$$E_{ihk} = \frac{\int P_i(k \mid \theta) f^{*f}(h - z \mid \theta) \varphi(\theta) \partial\theta}{\int f(h \mid \theta) \varphi(\theta) \partial\theta}$$

where $P_i(k|\theta)$ is the calculated probability that a person with θ gets an item score k on item i, $f(\square \mid \theta)$ is the conditional predicted test score distribution given θ, $f^{*i}(\square \mid \theta)$ represents the conditional predicted test score distribution without item i, and $\varphi(\theta)$ is the population distribution of θ. Thissen, Pommerich, Billeaud, and Williams (1995) developed the generalized recursive algorithm that can be used to compute $f(\square \mid \theta)$ and $f^{*i}(\square \mid \theta)$. The recursive algorithms needed to be implemented for S-X^2 are available through a SAS macro, IRTFIT (Bjorner, Smith, Stone, & Sun, 2007).

Kang and Chen (2008) found close to nominal Type I error rates in data generated to fit the RSM, PCM and GPCM for 5, 10 and 20-item tests and examinee sample sizes ranging from 500 to 5,000. They also found power estimates ranging from .57 to .98 when GPCM

and RSM were compared across all other conditions. Kang and Chen (2011) extended their previous study to the GRM, studying the effects of number of item score categories (3, 5), ability distribution (normal, uniform), size of the examinee sample (500, 1000, 2000), and test length (5, 10, 20). They found that with the exception of the condition with the longest test, smallest sample size, and largest score category, the Type I error rates ranged from .03 to .08, while power in detecting misfit due to multidimensionality and discrepancy from the GRM was adequate for large samples only.

Roberts (2008) generalized the $S\text{-}X^2$ to the generalized graded unfolding model (GGUM; Roberts, Donoghue, & Laughlin, 2000) for polytomous responses. The GGUM is a unidimensional polytomous proximity-based IRT model that assumes that examinees are more likely to receive higher item scores if they are located close to an item on the ability continuum. The $S\text{-}X^2$ and a corrected version $S\text{-}X_c^2$ performed best in terms of curtailing the Type I error (close to nominal rate), whereas power was highest with moderate to high misfit.

Glas and Falcón (2003) proposed another fit index based on a Lagrange multiplier test. As for the $S\text{-}X^{2,}$, examinees are grouped based on number-correct scores rather than trait scores, and the standard errors in item parameter estimates are taken into account. The result for this index was Type I error rates close to the nominal alpha level. This index is infrequently used given the lack of software available for its implementation.

Alternatives to Traditional Chi-Square Methods

Given the aforementioned issues with goodness-of-fit chi-square statistics relative to Type I error rates and power considerations, as well as lack of methods for visual investigation of fit, other researchers sought alternative methods of investigating polytomous model fit.

Model Selection. Related to investigations of model fit is the notion of model selection. Kang, Cohen and Sung (2009) examined the Akaike's information criterion (AIC; Akaike, 1974), the Bayesian information criterion (BIC; Schwartz, 1978), the deviance information index (DIC; Spiegelhalter, Best, Carlin, & Van der Linden, 2002), and the cross validation log likelihoods (CVLL) for the GRM, GPCM, PCM, and RSM and found that the BIC is the most accurate in selecting the correct polytomous model. However, fit statistics based on nested model comparisons are also somewhat sensitive to sample size. Moreover, there is no evidence that model selection statistics are appropriate when comparing models with different types of estimated parameters (i.e., adjacent category vs. cumulative boundary).

Nonparametric Approach to Testing Fit. Adequate Type I error rates and adequate power as well as graphical representations of misfit were established by Li and Wells (2006) and Liang and Wells (2009) for the polytomous case using a comparison between nonparametric and parametric IRFs first introduced by Douglas and Cohen (2001). The argument here is that the nonparametric approach imposes fewer restrictions on the shape of the IRF, and it is possible to conclude that the parametrically based model is incorrect if it differs substantially from the nonparametric IRF. The method involves estimating IRFs nonparametrically, finding the best fitting IRF for the parametric model of interest, and testing whether the distance between the two IRFs is significantly different. This difference can be described using the root integrated squared error (RISE), here specified for the GPCM:

$$RISE_i = \sqrt{\frac{\sum_{q=1}^{Q}\left(\frac{\sum_{q=1}^{Q}(P_{qk} - \hat{P}_{qk}^{non})^2}{Q}\right)}{K-1}}$$

where p_{qk} and $\hat{P}_{qk}^{non}$ are points on the IRF corresponding to the evaluation points for the model-based and nonparametric estimation methods for step IRF *k*, respectively, *Q* is the number of evaluation points used to obtain the kernel-smoothed IRF, and *K* is the total number of categories. The significance test for RISE may be derived using a parametric bootstrapping procedure where the proportion of RISEs from simulated data greater than the observed RISE value gives the approximated *p*-value for item *i*.

Li and Wells (2006) tested the nonparametric approach with the GRM across test lengths and sample sizes, concluding that the fit statistic performed well in terms of Type I error (close to nominal) and power (high). Liang and Wells (2009) found similar results for the GPCM. The advantages of this approach is that the Type I error rate was controlled and power was acceptable regardless of sample size and that graphical display of possible misfit is available. FORTRAN code was developed to implement the nonparametric approach to assessing fit and generating item responses (available from authors).

Posterior Predictive Checks. Rubin (1984) used posterior predictive model checking (PPC) to compare features of simulated data against observed data. Specifically, this method is based on the argument that if a model is a good fit for the data, then future data simulated from the model should be similar to the current data (Gelman, Carlin, Stern, & Rubin, 2003).

The posterior distribution for parameters is obtained by:

$$p(\boldsymbol{\delta} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\delta})\, p(\boldsymbol{\delta})$$

When $\mathbf{y}^{rep}$ is a matrix of replicated observations given the observed sample data, the probability distribution for future observations is:

$$p(\mathbf{y}^{rep} \mid \mathbf{y}) = \int p(\mathbf{y}^{rep}, \boldsymbol{\delta})\, p(\boldsymbol{\delta} \mid \mathbf{y}) d\boldsymbol{\delta},$$

To evaluate misfit, or the difference between the observed and predicted data is summarized in discrepancy measures. Then, posterior predictive *p*-values, or summaries of relative occurrence of the value for an observed discrepancy measure in the distribution of discrepancy values from replicated data, are used to summarize model fit (Zhu & Stone, 2011). When $F(\mathbf{y})$ is a function of the data, and $F(\mathbf{y}^{\text{rep}})$ is the same function, but applied to replicated data, the Bayesian *p*-value is:

$$p - value = p[F(\mathbf{y}^{rep}) \geq F(\mathbf{y}) \mid \mathbf{y}]$$

The *p*-value can be interpreted as the proportion of replicated data sets for which function values $F(\mathbf{y}^{rep})$ are greater than or equal that of the function $F(\mathbf{y})$. The PPC *p*-values can be interpreted such that the value of .5 describes no systematic differences between observed and predictive discrepancy measures (i.e., adequate fit) and values close to 0 or 1 indicate that the observed discrepancies do not agree with the posterior predictive discrepancy measures (i.e., misfit) (Zhu & Stone, 2011).

Discrepancy measures can be any statistic that has the potential to reveal sources of misfit, including violations of the IRT assumption of local item independence, or dimensionality. A chi-square statistic can be used to detect differences between observed and expected score frequencies. At the item level, any of the chi-square fit indices described above (e.g., $S\text{-}X^2$) could be used as a discrepancy measure.

Zhu and Stone (2011) used PPCs to evaluate misfit to a GRM due to multidimensionality and local dependence. Specifically, they found that pairwise measures (specifically, global odds ratio and Yen's Q_3) were successful in detecting violations of unidimensionality and local independence assumptions, and Q_3 exhibited greatest empirical power. Item-fit statistic Q_1 was once more found to be ineffective in identifying misfit.

WinBUGS (Spiegelhalter, Thomas, Best, & Lunn, 2003) software was used in estimating the GRM with MCMC. However, SAS procedure MCMC could also be used for estimating the model.

The PPC method has several advantages for investigating IRT model fit. First, PPCs account for error in parameter estimation by using parameter posterior distributions instead of point estimates. Second, the researcher can construct an empirical null sampling distribution rather than a potentially inappropriate analytically derived distribution. Finally, model complexity is less of an issue for Bayesian methods.

It is clear that model fit evaluation can be conducted accurately using fairly sophisticated methods. Such methods should be considered for future studies, as software and further research regarding these methods become more available. In this project, the *S-*X^2 and PARSCALE's G^2 were used to evaluate item fit.

The CPA Exam utilizes a three-stage Computer Adaptive Multistage Testing (caMST) delivery method for the multiple choice question (MCQ) component of the Exam, which consists of pre-constructing content-balanced modules based on test information targets and item exposure controls. Regardless of the caMST MCQ routing, each candidate taking the CPA exam responds to a group of Task-Based Simulations (TBS) during the last portion of the exam (i.e., prior information about candidate ability is not available for TBS section scoring) for three of the four available exam sections: AUD, FAR and REG. Approximately twenty-four pre-assembled panels per section contain TBS items. Thus, because candidates are exposed to different combinations of panels during an exam situation, the full response data set for each testing window contains sparseness. It is important to evaluate whether polytomous model calibration with sparse data is possible and provides

similar information to calibration with the currently operational 3PL model and to that of smaller, complete, panel-specific datasets.

Method

Major goals of the project were to compare a) the GPCM and GRM models relative to statistical and theoretical fit, b) the rank order of ability estimates across models, and c) the shape of the test information functions (TIFs) between polytomous and dichotomous model calibrations. Two types of analyses were conducted to shed light on the viability of the testlet-based summation scoring method and subsequent polytomous model fit of the CPA Exam TBS items: 1) sparse data calibration with the GRM and the GPCM (and 1PL implementations of each); 2) individual panel calibration with the GRM and the GPCM (and 1PL implementations of each).

Sparse Data Analysis

Sparse data sets for TBS data were created by summing candidate responses to measurement opportunities (MOs) belonging to each common task stimulus while considering the full testing window dataset for each section (N_{AUD} = 16,326, N_{FAR} =17,672, N_{REG} =17,322). Summed TBS data were analyzed separately for each section using PARSCALE 4.1 (Muraki & Bock, 2003). In addition, BILOG-MG 3.0 was used to fit the 3PL model to the original binary data. A not-presented key was developed for each section data set to distinguish between responses that earned zero credit and missingness due to the item not being included on the candidate's form.

The AUD, FAR, and REG data sets were analyzed using the GPCM, PCM, GRM and the GRM constrained to have one common slope across all TBS (1PL-GRM). In addition, the 3PL model was fit for comparison with PIRT models of interest. Operational priors and

calibration/estimation settings in BILOG-MG were used for the 3PL, and similar calibration/estimation settings were applied to summated TBS scores.

For all models, calibration was conducted using the natural logistic function, 30 quadrature points, 100 E-M cycles (25 for PCM and 1PL GRM), 500 Newton cycles (5 for PCM and 1PL GRM), and the default E-M cycles criterion of .001 for convergence. Scale score estimation was conducted using the expected a priori (EAP) method. The original scores were rescaled to have the mean of 0 and standard deviation of 1. If the model did not successfully converge with default settings, several methods were employed to facilitate convergence. First, the number of E-M and Newton cycles were reset to a higher value to affect convergence. Second, the convergence criterion for the E-M cycles was reset to a larger value (no higher than .01). If increasing the number of estimation cycles and the convergence criterion still resulted in item parameter calibration stage terminating and an error message produced an indication of the problematic TBS/category within the TBS, the location parameter was gradually adjusted using the "CADJUST" option in PARSCALE until errors associated with the TBS or a specific response category were eliminated. The maximum value of the location adjustment was .2 and typically did not exceed .10.

Generally, the same TBSs caused convergence issues across different panels. In the AUD section, eight out of 35 TBSs were the cause of convergence issues in ten panels for the GRM calibration and four panels in the GPCM calibration. In the FAR section, eight out of 31 TBSs caused issues in eight panels for the GRM calibration and four panels for the GPCM calibration. In the REG section, two out of 27 TBSs caused issues in two panels during the GRM calibration and four panels during the GPCM calibration. In the case of the GRM calibration, convergence issues were caused by low frequency of responses at the tails of the

score distribution, whereas in the case of the GPCM the cause tended to be either unknown or related to an irregular score distribution. In order to fit the GRM to the AUD section sparse data, response categories with low response frequencies were collapsed. Specifically, TBSs associated with error messages regarding response frequencies that did not occur in a descending order were flagged and checked for low response counts at the extremes of the score distribution. Then, response categories making up less than .1 percent of the score distribution were collapsed with an adjacent score category.

Panel Data Analysis

Within each section, MOs belonging to common TBSs were summed together and assigned to their original panels. The AUD, FAR, and REG panel data sets were analyzed using the GPCM, PCM, GRM and the 1PL GRM. In addition, the 3PL model was fit for comparison using operational priors and calibration/estimation settings in BILOG-MG. Each TBS (i.e., the summation of the relevant MOs) was considered separately during the item parameter calibration and theta estimation. Initial settings and methods of ensuring convergence were similar to those for the sparse data analysis.

Results

Descriptive Information

Frequencies of TBSs with specific number of categories in each exam section are displayed in Figure 1. Of the total existing 97 TBSs, 19 had two possible response options (i.e., one measurement opportunity) across the three sections. Because TBSs were reused on different panels, the number of dichotomous TBS *inclusions* was 75 out of 403 (N_{AUD} = 29, N_{FAR} = 24, N_{REG} = 22). That is, the 19 dichotomous TBSs were included in more than one panel, resulting in multiple instances of the TBS. The most frequent numbers of response categories were six in the AUD section, nine in the FAR section, and seven in the REG

section. Parameters for TBSs with two response categories were calibrated using the 2PL DIRT model when included in the GRM and GPCM calibrations, and the 1PL DIRT model when included in the 1PL GRM and PCM calibrations.

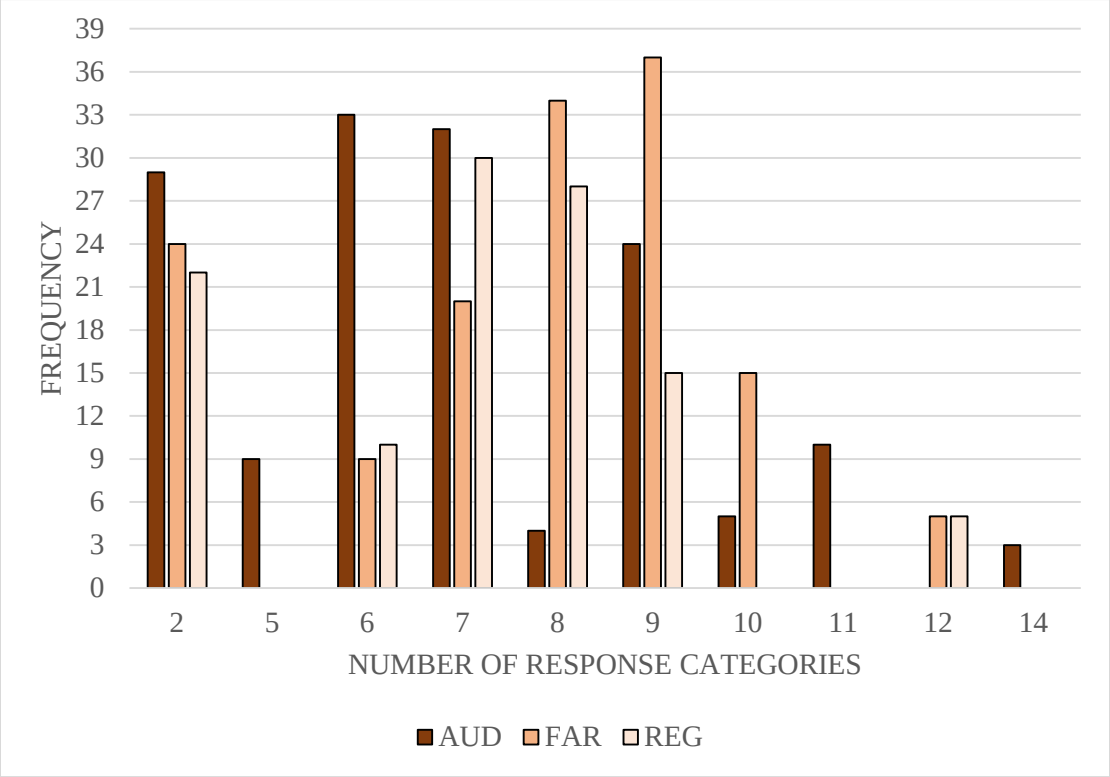


Figure 1. TBS frequencies at response category number.

Comparison of Theta Estimates

Ability estimates were compared between the sparse and full panel data sets for all models. Theta values at ten percentiles are presented in Tables 1, 2, and 3 for the AUD, FAR, and REG sections, respectively. The 1PL version of the GRM in the sparse case did not converge for any of the sections, suggesting that a common slope parameter was overly restrictive. Generally, ability estimates were very similar between the full and sparse data sets. Some departures occurred in the extremes of the ability distribution whereby the sparse

data sets tended to yield lower ability estimates in the 3PL case, and higher ability estimates in the PIRT model case.

Spearman's rank-order correlation was calculated for the ability estimates yielded by the 3PL, GRM and GPCM models across the sparse and panel datasets for all three sections. Scatterplots of sparse and full panel ability estimates are presented in Figures 2 through 9. Agreement between sparse and panel ability estimates was highest in the polytomous case, although all rank-order correlations were generally high, ranging between .978 (3PL, AUD) and .995 (GRM, FAR). Combined with the information provided by the percentile break-down of ability estimates, the comparison between sparse and panel results reflect a close agreement between estimates from the same model, suggesting that sparsity is not causing problems with estimation of ability. As can be observed from Tables 4 through 6, agreement between the polytomous models was generally slightly higher. Overall, ability estimates were "overestimated" by the polytomous models when compared to the 3PL. Interestingly, rank-order correlations were systematically highest between the 3PL and the PCM ability estimates. However, claims about the absolute agreement between estimates cannot be made based on this relationship. That is, agreement regarding the actual value of theta cannot be described using a rank-order coefficient.

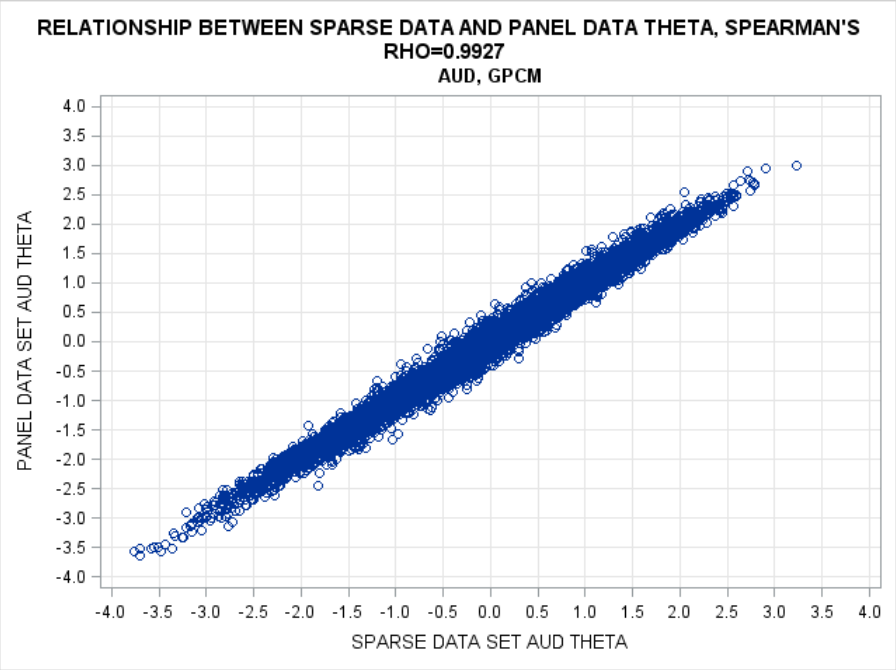


Figure 2. Scatterplot depicting AUD section variability of sparse ability estimates and panel ability estimates for the GPCM. Theta estimates were scaled such that the mean ability was 0 and standard deviation was 1.

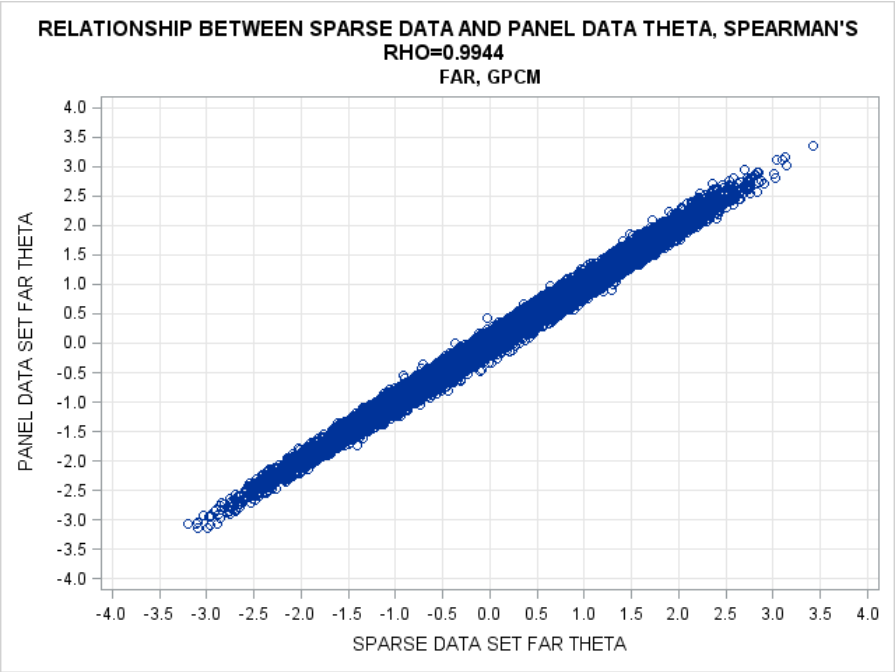


Figure 3. Scatterplot depicting FAR section variability of sparse ability estimates and panel ability estimates for the GPCM. Theta estimates were scaled such that the mean ability was 0 and standard deviation was 1.

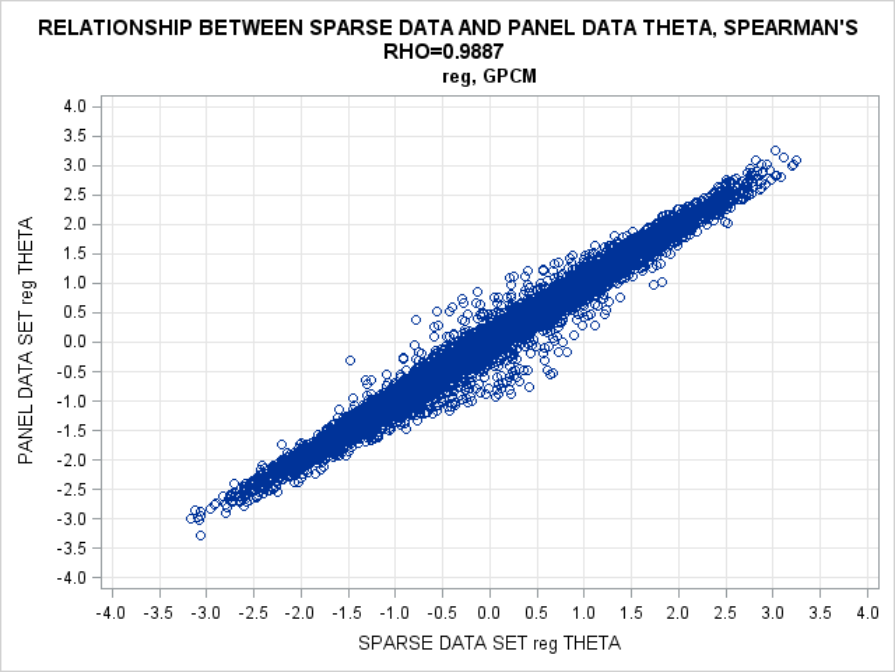


Figure 4. Scatterplot depicting REG section variability of sparse ability estimates and panel ability estimates for the GPCM. Theta estimates were scaled such that the mean ability was 0 and standard deviation was 1.

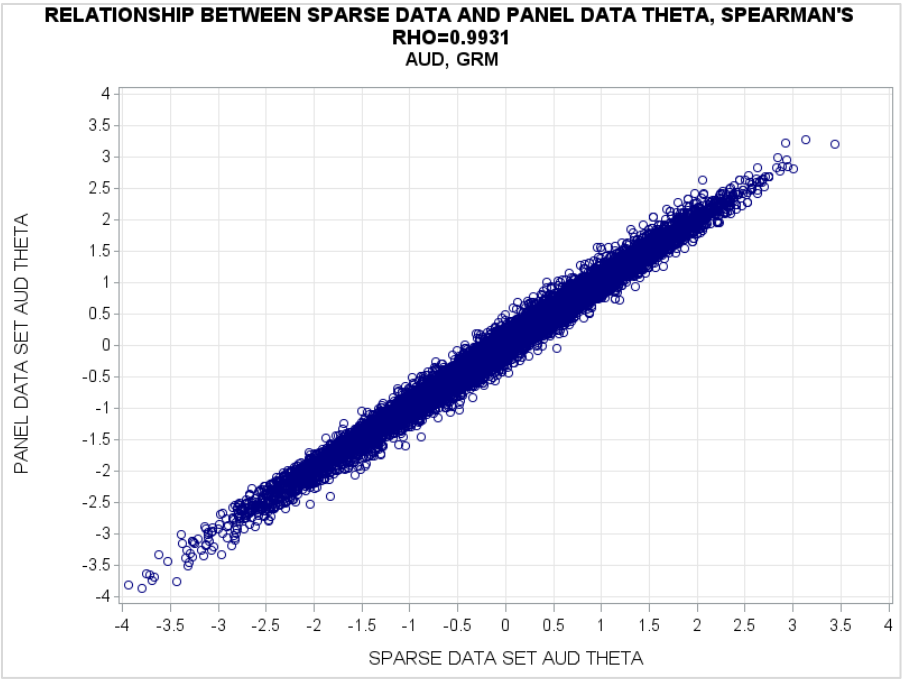


Figure 5. Scatterplot depicting AUD section variability of sparse ability estimates and panel ability estimates for the GRM. Theta estimates were scaled such that the mean ability was 0 and standard deviation was 1.

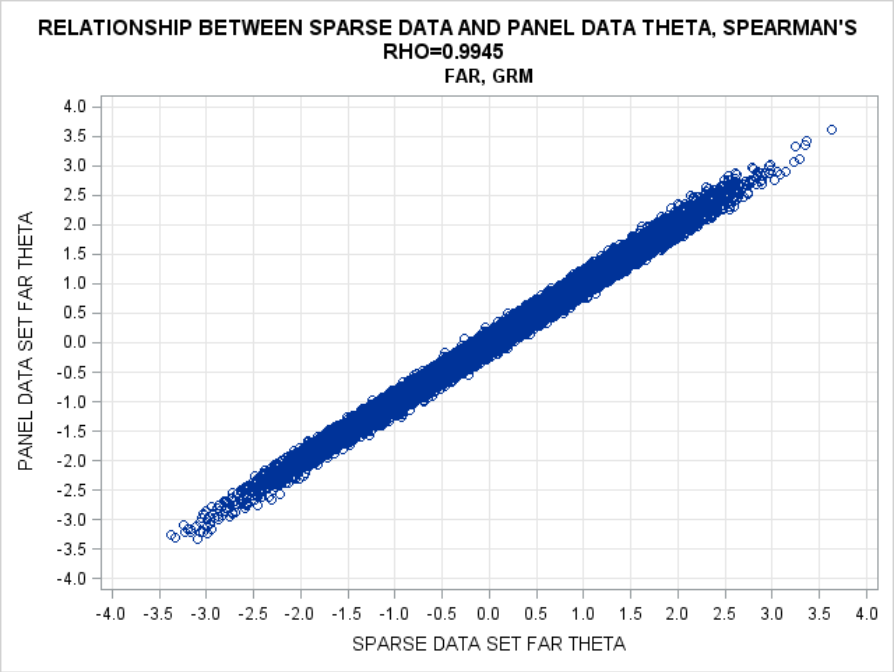


Figure 6. Scatterplot depicting FAR section variability of sparse ability estimates and panel ability estimates for the GRM. Theta estimates were scaled such that the mean ability was 0 and standard deviation was 1.

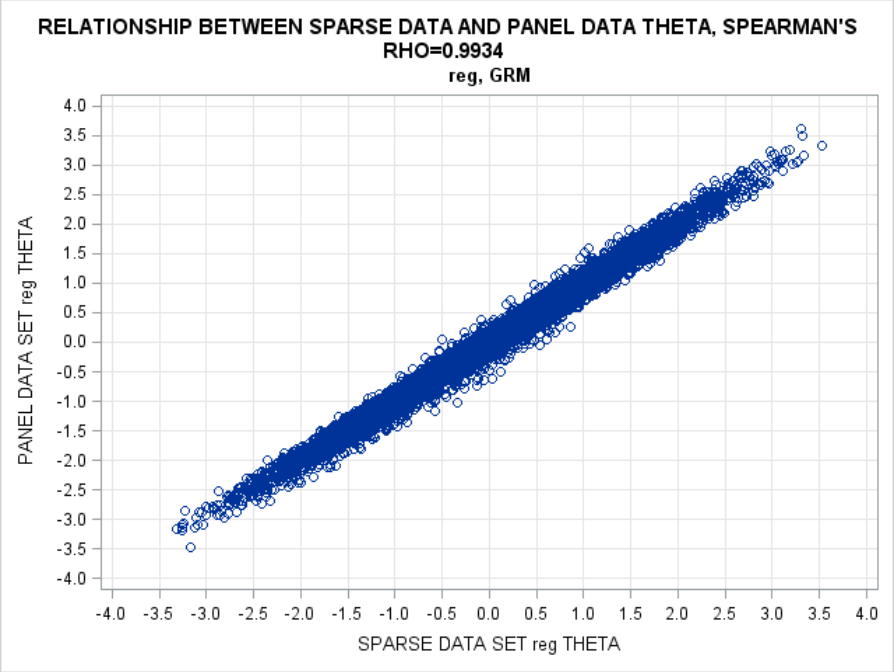


Figure 7. Scatterplot depicting REG section variability of sparse ability estimates and panel ability estimates for the GRM. Theta estimates were scaled such that the mean ability was 0 and standard deviation was 1.

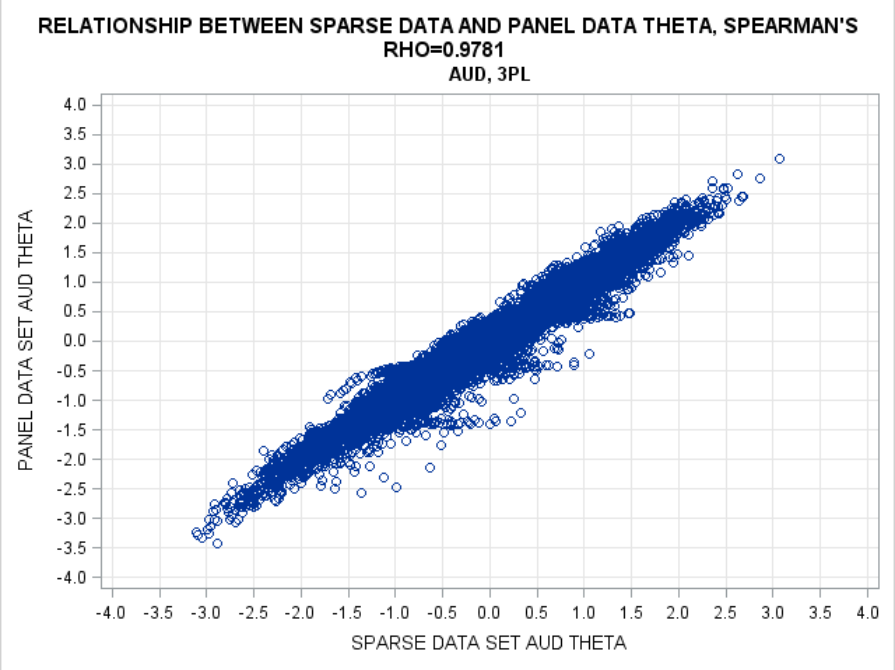


Figure 8. Scatterplot depicting AUD section variability of sparse ability estimates and panel ability estimates for the 3PL model. Theta estimates were scaled such that the mean ability was 0 and standard deviation was 1.

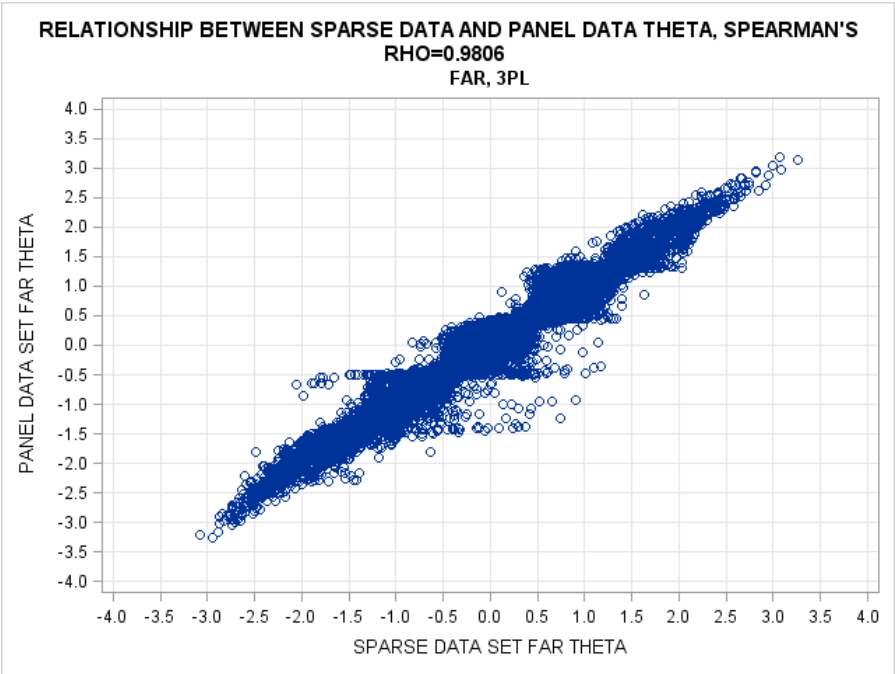


Figure 9. Scatterplot depicting FAR section variability of sparse ability estimates and panel ability estimates for the 3PL model. Theta estimates were scaled such that the mean ability was 0 and standard deviation was 1.

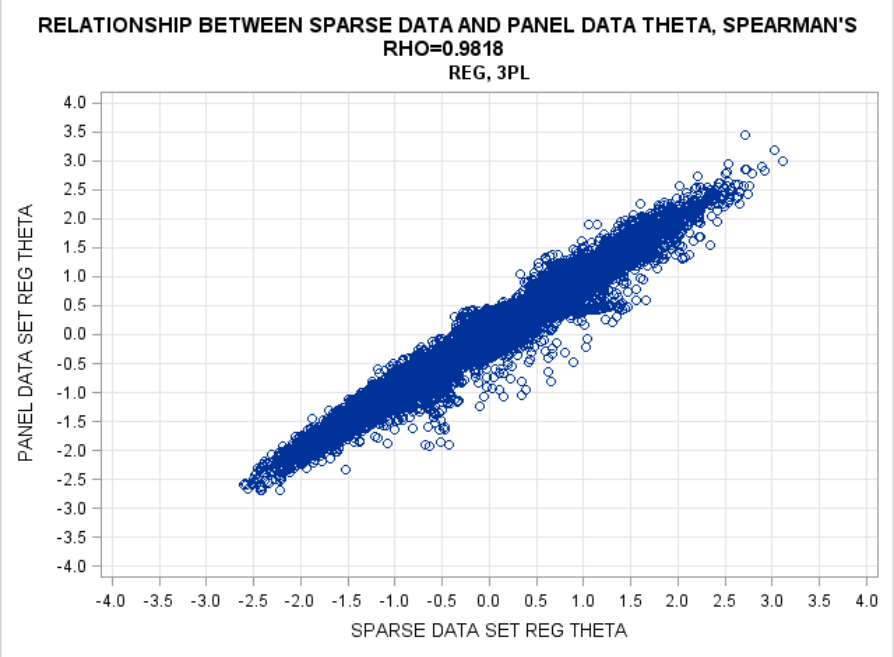


Figure 10. Scatterplot depicting REG section variability of sparse ability estimates and panel ability estimates for the 3PL model. Theta estimates were scaled such that the mean ability was 0 and standard deviation was 1.

Convergence Rates

Some analyses did not reach convergence. Specifically, in the AUD panel case (24 panels), five panels did not converge with the 1PL-GRM, and nine panels did not converge with the PCM. In the FAR panel case (24 panels), eight panels did not converge with the PCM, and nine panels did not converge with the 1PL GRM. In the REG panel case (23 panels), one panel did not converge with the GPCM, 14 panels did not converge with the 1PL-GRM, and two panels did not converge with the PCM. Upon further examination, panels that contained TBS with a large number of categories (> 9) typically returned a phase 2 (item calibration) PARSCALE error message “Initial category parameters must be in

descending order" during 1PL GRM calibration, and "Matrix is singular" with no further explanation during PCM calibration.

Information

The 3PL, GRM, and GPCM test information functions (TIFs) were obtained from the panel item parameters and ability estimates. For PIRT models, PARSCALE output the logistic item information function for the graded or partial credit models as proposed by Samejima (1974). The dichotomous item information is a simplification of the partial credit model item information to the dichotomous case. To obtain the polytomous TIFs, polytomous TBS information was summed across all available TBSs. To obtain the dichotomous TIFs, individual measurement opportunity information was summed. Due to the finding that sparse and panel calibrations provide essentially the same information, Figures 10, 11, and 12 display GRM, GPCM, PCM and 3PL TIFs for the overall sparse datasets in the AUD, FAR, and REG sections, respectively. Information resulting from the PCM and GPCM calibrations was approximately similar. Thus, graphs of the GRM, GPCM, and 3PL information functions for each panel are displayed in Appendix A.

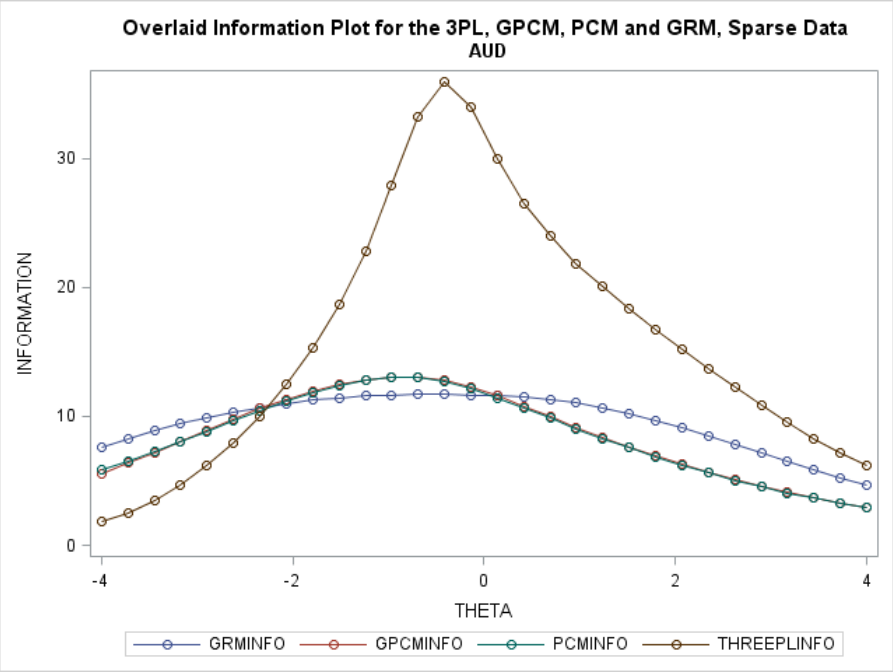


Figure 10. Test Information Functions for the AUD section. TIFs are displayed for the GRM (blue), GPCM(red), PCM(green) and 3PL(brown).

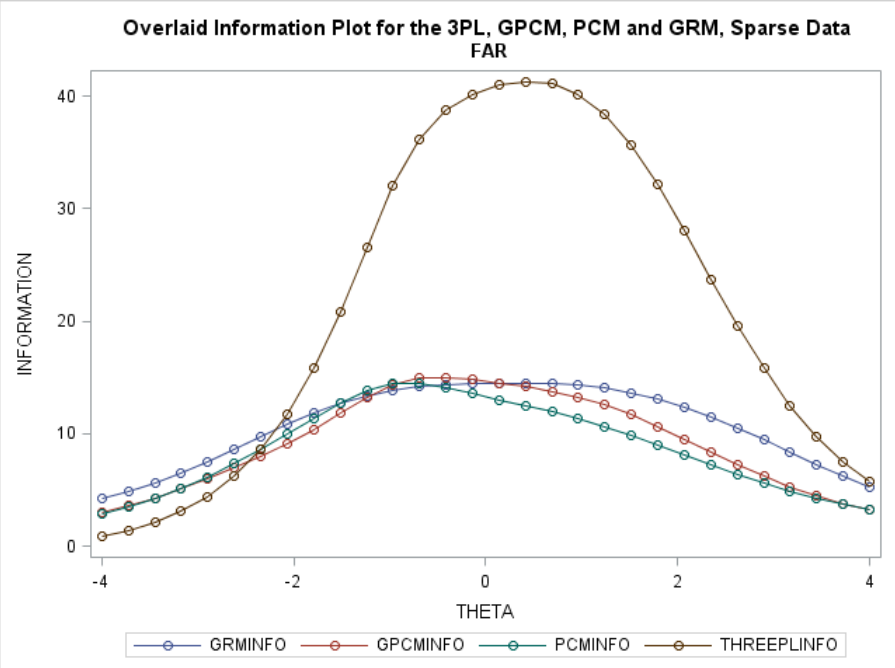


Figure 11. Test Information Functions for the FAR section. TIFs are displayed for the GRM (blue), GPCM(red), PCM(green) and 3PL(brown).

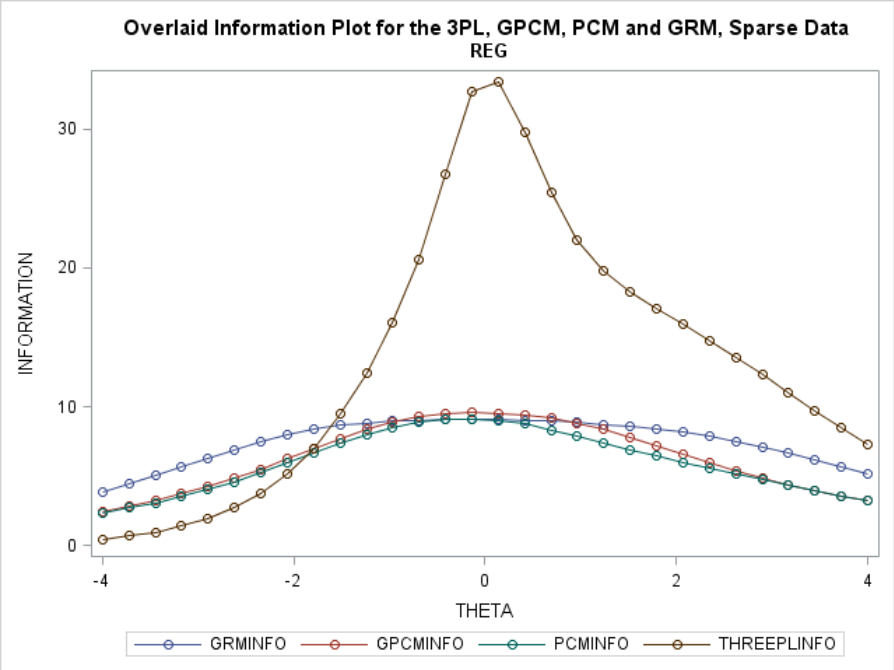


Figure 12. Test Information Functions for the REG section. TIFs are displayed for the GRM (blue), GPCM (red), PCM (green) and 3PL (brown).

In general, the information curves showed the expected pattern of relatively low information displayed by the PIRT functions as compared with the 3PL. Compared with the GPCM, the GRM test information function was more evenly spread across the latent ability continuum, providing more information in the extremes. The GPCM function tended to have a peaked quality with less information around the tails of the distribution. It should be noted that for an accurate comparison among information curves from different models, it is necessary to equate the item parameters. Once equated, information from both models could be plotted on the same scale, and thus, compared more fairly. Current results should be viewed as a rough approximation of information comparison. To supplement the information curves, Table 7 displays descriptive statistics regarding the standard error of measurement associated with ability estimates acquired with each model. Naturally, the model with the smallest number of estimated parameters has the smallest error (3PL).

Table 1

AUD Section Quantile Latent Ability Estimates across Models (Sparse and Panel)

Quantile	GRM*	SGRM	1PL GRM	S1PL GRM	GPCM*	SGPCM	PCM	SPCM	3PL	S3PL
100% Max	3.259	3.440	3.382	N/A	3.002	3.229	3.073	3.298	3.085	3.062
99%	2.134	2.138	2.138	N/A	2.082	2.081	2.055	2.044	1.974	1.934
95%	1.590	1.582	1.589	N/A	1.582	1.571	1.565	1.551	1.474	1.448
90%	1.272	1.278	1.260	N/A	1.269	1.276	1.254	1.266	1.199	1.175
75% Q3	0.700	0.691	0.692	N/A	0.720	0.709	0.690	0.716	0.602	0.633
50% Median	0.043	0.041	0.039	N/A	0.042	0.044	0.040	0.045	0.092	0.067
25% Q1	-0.655	-0.652	-0.649	N/A	-0.669	-0.656	-0.663	-0.660	-0.567	-0.582
10%	-1.317	-1.320	-1.307	N/A	-1.331	-1.330	-1.329	-1.329	-1.286	-1.265
5%	-1.716	-1.716	-1.722	N/A	-1.722	-1.714	-1.707	-1.706	-1.574	-1.559
1%	-2.473	-2.455	-2.509	N/A	-2.453	-2.460	-2.456	-2.462	-2.274	-2.245
0% Min	-3.876	-3.935	-3.866	N/A	-3.636	-3.755	-3.700	-3.782	-3.423	-3.112

Note. *The location of an item was adjusted to assist convergence. "S" indicates a model that was fit using the sparse data set.

Table 2

FAR Section Quantile Latent Ability Estimates across Models (Sparse and Panel)

Quantile	GRM*	SGRM*	1PL GRM	S1PL GRM	GPCM*	SGPCM*	PCM	SPCM	3PL	S3PL
100% Max	3.621	3.637	3.593	N/A	3.358	3.421	3.179	3.394	3.191	3.258
99%	2.335	2.334	2.321	N/A	2.310	2.294	2.251	2.299	2.183	2.145
95%	1.650	1.658	1.656	N/A	1.667	1.669	1.698	1.678	1.412	1.450
90%	1.297	1.302	1.290	N/A	1.314	1.319	1.315	1.320	1.288	1.271
75% Q3	0.669	0.675	0.662	N/A	0.674	0.679	0.673	0.688	0.528	0.562
50% Median	0.006	-0.002	0.001	N/A	0.001	-0.006	0.000	-0.008	0.105	0.078
25% Q1	-0.681	-0.679	-0.681	N/A	-0.695	-0.687	-0.690	-0.689	-0.540	-0.553
10%	-1.286	-1.290	-1.273	N/A	-1.291	-1.285	-1.303	-1.278	-1.346	-1.318
5%	-1.646	-1.653	-1.662	N/A	-1.654	-1.650	-1.640	-1.649	-1.641	-1.649
1%	-2.303	-2.321	-2.311	N/A	-2.279	-2.299	-2.284	-2.294	-2.299	-2.295
0% Min	-3.322	-3.375	-3.605	N/A	-3.144	-3.196	-3.138	-3.231	-3.271	-3.075

Note. *The location of an item was adjusted to assist convergence. "S" indicates a model that was fit using the sparse data set.

Table 3

REG Section Quantile Latent Ability Estimates across Models (Sparse and Panel)

Quantile	GRM*	SGRM*	1PL GRM	S1PL GRM	GPCM*	SGPCM*	PCM	SPCM	3PL	S3PL
100% Max	3.617	3.521	3.535	N/A	3.259	3.237	3.285	3.296	3.455	3.106
99%	2.311	2.333	2.320	N/A	2.270	2.305	2.283	2.303	2.077	2.018
95%	1.646	1.642	1.630	N/A	1.650	1.653	1.647	1.660	1.447	1.445
90%	1.284	1.283	1.259	N/A	1.283	1.291	1.284	1.283	1.222	1.188
75% Q3	0.676	0.674	0.669	N/A	0.691	0.685	0.670	0.702	0.582	0.605
50% Median	0.007	0.001	0.019	N/A	0.010	0.004	0.014	-0.007	0.076	0.056
25% Q1	-0.674	-0.668	-0.655	N/A	-0.694	-0.682	-0.692	-0.673	-0.659	-0.643
10%	-1.293	-1.280	-1.272	N/A	-1.302	-1.295	-1.300	-1.298	-1.226	-1.226
5%	-1.652	-1.660	-1.673	N/A	-1.650	-1.654	-1.662	-1.669	-1.517	-1.514
1%	-2.335	-2.372	-2.432	N/A	-2.277	-2.296	-2.304	-2.323	-2.065	-2.031
0% Min	-3.472	-3.321	-3.450	N/A	-3.291	-3.166	-3.249	-3.222	-2.683	-2.609

Note. *The location of an item was adjusted to assist convergence. "S" indicates a model that was fit using the sparse data set.

Table 4

AUD Section Spearman Correlations of Available Model Ability Estimates

	3PL(S)	3PL	GPCM(S)	GPCM	PCM(S)	PCM	GRM(S)	GRM
3PL	**0.98**							
GPCM(S)	0.94	0.90						
GPCM	0.94	0.90	**0.99**					
PCM(S)	0.95	0.91	0.99	0.98				
PCM	0.95	0.91	0.98	0.98	**0.99**			
GRM(S)	0.93	0.89	0.99	0.99	0.98	0.98		
GRM	0.92	0.89	0.99	0.99	0.98	0.97	**0.99**	
GRM1PL	0.89	0.85	0.97	0.97	0.97	0.97	0.98	0.98

Note. N = 7482. Bolded values indicate the Spearman correlation between the panel and sparse ability estimates for the same model.

Table 5

FAR Section Spearman Correlations of Available Model Ability Estimates

	3PL(S)	3PL	GPCM(S)	GPCM	PCM(S)	PCM	GRM(S)	GRM
3PL	**0.97**							
GPCM(S)	0.96	0.92						
GPCM	0.96	0.92	**0.99***					
PCM(S)	0.98	0.94	0.99	0.98				
PCM	0.98	0.94	0.98	0.98	**0.99***			
GRM(S)	0.95	0.90	0.99	0.99	0.98	0.97		
GRM	0.94	0.90	0.99	0.99	0.97	0.97	**0.99***	
GRM1PL	0.93	0.89	0.97	0.97	0.97	0.98	0.98	0.98

Note. N = 7347. Bolded values indicate the Spearman correlation between the panel and sparse ability estimates for the same model. *Correlation is higher than .99.

Table 6

REG Section Spearman Correlations of Available Model Ability Estimates

	3PL(S)	3PL	GPCM(S)	GPCM	PCM(S)	PCM	GRM(S)	GRM
3PL	**0.98**							
GPCM(S)	0.95	0.92						
GPCM	0.94	0.92	**0.99***					
PCM(S)	0.97	0.94	0.99	0.98				
PCM	0.96	0.95	0.98	0.98	**0.99***			
GRM(S)	0.95	0.92	0.99	0.99	0.98	0.98		
GRM	0.93	0.91	0.99	0.99	0.98	0.98	**0.99***	
GRM1PL	0.92	0.90	0.97	0.98	0.97	0.98	0.98	0.99

Note. N = 5839. Bolded values indicate the Spearman correlation between the panel and sparse ability estimates for the same model. *Correlation is higher than .99.

Table 7

Standard Errors of Latent Ability Estimates across Considered Models

Model	N	Mean	SD	MIN	MAX
AUD Section					
PCM (S)	16326	0.74	0.04	0.65	0.89
GPCM (S)	16326	0.73	0.05	0.59	0.92
GRM(S)	16326	0.72	0.06	0.55	1.06
3PL(S)	16326	0.44	0.08	0.11	0.82
PCM	10228	0.77	0.09	0.6	0.99
GPCM	14940	0.71	0.08	0.57	0.99
GRM	16326	0.71	0.08	0.54	1.12
1PL GRM	16326	0.71	0.08	0.54	1.12
3PL	15671	0.41	0.11	0.02	0.89
FAR Section					
PCM (S)	17672	0.64	0.04	0.54	0.78
GPCM (S)	17672	0.62	0.05	0.51	0.8
GRM(S)	17672	0.62	0.06	0.48	0.92
3PL(S)	17672	0.36	0.1	0.09	0.76
PCM	11781	1.56	1.96	0.49	6.61
GPCM	17672	0.62	0.07	0.49	0.85
GRM	17672	0.61	0.07	0.45	1
1PL GRM	17672	0.61	0.07	0.45	1
3PL	17672	0.34	0.12	0.01	1.42
REG Section					
PCM (S)	17322	0.78	0.05	0.67	0.96
GPCM (S)	17322	0.77	0.06	0.65	0.96
GRM(S)	17322	0.77	0.07	0.61	1.1
3PL(S)	17322	0.46	0.11	0.07	0.75
PCM	15697	1.05	1.09	0.6	5.99
GPCM	16555	0.75	0.09	0.59	1
1PL GRM	7464	0.76	0.08	0.58	1.12
GRM	17322	0.77	0.1	0.55	1.16
3PL	17322	0.43	0.13	0.01	1.1

Note. (S) indicates that the model was fit to a sparse data set.

Model Fit

PIRT model fit for panel data was evaluated using PARSCALE's G^2 statistic and IRTPro 2.1's (Cai, Thissen, & du Toit, 2011) $S\text{-}X^2$. In each panel and section, the G^2 chi-square test statistic and *p*-value were recorded. Overall 1,377 G^2 statistics were produced for four models across the three sections. To summarize the fit information, *p*-values associated with the chi-square tests of significance were categorized such that *p*-values less than .05 were arbitrarily set to indicate less than adequate fit. GRM, GPCM, 1PL GRM, and PCM $S\text{-}X^2$ statistics for all TBSs in each section are listed in Appendix B. As a reminder, $S\text{-}X^2$ statistics use adjusted degrees of freedom and thus constitute a better estimate of observed to expected fit ratios (Kang & Chen, 2007; 2011).

As summarized in Table 8, GRM fit was associated with G^2 *p*-values larger than .05 most frequently across all sections. Specifically, the GRM displayed adequate fit 65% of the time in the AUD section, 63% of the time in the FAR section, and 43% of the time in the REG section. In contrast, when compared with the GPCM, PCM, and the 1PL GRM, $S\text{-}X^2$ statistics associated with the GRM did not consistently indicate superior fit. Overall, panel data fit appeared equivalent between the GPCM and the GRM.

Table 8

Summary of Fit Information: Frequency and Percent of G^2 and S-X^2 Chi-Squared Statistics

		G^2				$S\text{-}X^2$			
	Significance Level	GPCM(%)	GRM(%)	PCM(%)	GRM1(%)	GPCM(%)	GRM(%)	PCM(%)	GRM1(%)
AUD	$p > .05$	80(54.1)	97(65.1)	43(46.7)	60(51.3)	133(89.3)	127(85.2)	125(83.9)	117(78.5)
	$.01 < p \leq .05$	25(16.9)	16(10.7)	10(10.9)	14(12.0)	10(6.7)	17(11.4)	16(10.7)	19(12.8)
	$.001 < p \leq .01$	13(8.8)	9(6.0)	11(12.0)	14(12.0)	3(2.0)	4(2.7)	6(4.0)	6(4.0)
	$p \leq .001$	30(20.3)	27(18.1)	28(30.4)	29(24.8)	3(2.0)	1(0.7)	2(1.3)	7(4.7)
FAR	$p > .05$	72(50.0)	90(62.5)	32(31.4)	47(49.0)	122(84.7)	126(87.5)	110(76.4)	106(73.6)
	$.01 < p \leq .05$	23(16.0)	18(12.5)	10(9.8)	12(12.5)	14(9.7)	10(6.9)	18(12.5)	12(8.3)
	$.001 < p \leq .01$	18(12.5)	16(11.1)	9(8.8)	9(9.4)	6(4.2)	8(5.6)	9(6.3)	14(9.7)
	$p \leq .001$	31(21.5)	20(13.9)	51(50.0)	28(29.2)	2(1.4)	0(0.0)	7(4.9)	12(8.3)
REG	$p > .05$	41(37.3)	49(42.6)	32(30.5)	18(36.0)	104(90.4)	104(90.4)	94(81.7)	90(78.3)
	$.01 < p \leq .05$	16(14.6)	18(15.7)	7(6.7)	7(14.0)	7(6.1)	6(5.2)	15(13.0)	13(11.3)
	$.001 < p \leq .01$	8(7.3)	11(9.6)	15(14.3)	6(12.0)	1(0.9)	2(1.7)	2(1.7)	4(3.5)
	$p \leq .001$	45(40.9)	37(32.2)	51(48.6)	19(38.0)	3(2.6)	3(2.6)	4(3.5)	8(7.0)

Note. Percentages do not add to 100 due to rounding.

Discussion

In this study, the impact of an alternative scoring process for the Uniform CPA Exam TBSs was explored. In theory, the use of performance-based, "innovative" items should increase the fidelity between the test content and performance in practice (Scalise et al., 2007), yielding more accurate estimates of candidates' ability. Moreover, polytomous models have the potential to alleviate the problem of local item dependency by considering dichotomously scored items as a summed polytomous response. The GRM (Samejima, 1969) and the GPCM (Muraki, 1992) were considered as two theoretically appropriate models for use with the Uniform CPA Exam task-based simulation data sets.

In the past, it has been suggested that the theoretical choice between the GPCM and GRM is somewhat arbitrary (e.g., Ostini & Nering, 2005). Essentially, the only difference between the two models is purely mathematical. Therefore, the current project looked at model fit, information and convergence rates to evaluate the feasibility of using PIRT models for the scoring of the CPA Exam TBS. The examination of item fit statistics revealed essentially equivalent fit of both models to TBS response data. When roughly compared with the GPCM information functions, the GRM provided information over a wider span of latent ability estimates.

Several instances of item calibration did not reach an admissible solution. Overwhelmingly, convergence issues occurred during 1PL calibrations, with both PCM and 1PL-GRM returning errors and iteration termination. It is expected that the requirement of a descending order of categories for the 1PL GRM was one cause of convergence issues. The graded response family of models requires that the *b* parameters are ordered, while the PCM and GPCM do not. When the number of categories within a TBS was relatively large (≥ 9), the panel frequencies were low for extreme categories, which precluded the fine differentiation of order,

ultimately impacting the calibration. Also, the sample size associated with each panel did not exceed N = 800, which may have contributed to calibration difficulties with models requiring a common slope across all items.

Data sparseness had a substantial effect on convergence rates, but not on theta estimates. The 1PL GRM did not yield successful convergence in any of the exam sections when the sparse data set was considered. Nevertheless, for the PIRT models that did converge, panel-based and sparse data set latent ability estimates were highly correlated (> .99).

It is important to note that information functions as produced in this study should be considered only for rough comparisons among the models. The three studied functions cannot be directly compared due to the differing calculations used to obtain each model's function. Comparisons between the two PIRT models could be theoretically problematic as discrimination parameters contribute to the information function differently. Whereas the GRM algebraic formula for the *a*-parameter remains the same for any number of response categories, the GPCM *a*-parameter values artificially decrease with an increase in the number of response categories (Yurekli, 2010). Because of the consistency of the *a*-parameter calculation for any number of response categories, it has been suggested that the GRM is more appropriate for use with ordered response data (Jansen & Roskam, 1986; Yurekli, 2010). However, the close relationships between the GRM and GPCM ability estimates observed in this study may suggest that information functions could be roughly comparable. Future research should focus on obtaining more comparable information functions between the dichotomous 3PL and polytomous models.

In order to obtain a fair comparison of information, it is necessary to equate the information functions (Fitzpatrick & Dodd, 1997). Popular equating methodologies for this purpose include true score equating based on test characteristic curves (TCC) (Stocking & Lord,

1983). Another recommended method for equating when mathematically different models have been used in calibration is presented by Fitzpatrick and Dodd (1997) in a conference presentation. However, thorough research regarding appropriate ways to equate the two models prior to obtaining the information function is largely incomplete and scarce (Dodd, 2014, personal communication). For example, Jiao, Liu, Haynie, Woo, and Gorham (2011) provided information comparisons between the 2PL and the PCM, although the methodology was not explicit.

In future research, the shape of the information functions should be explored for fruitful comparisons between the 3PL and PIRT models. It is known that interdependent items have the effect of reducing test length if items are redundant (e.g., Sireci & Thissen, 1991). Whereas the 3PL model assumes that MOs within a TBS are independent, it may be the case that interdependencies between MOs exist in that examinees jointly succeed on certain tasks, which reduces the assessment's true reliability. In this study the visual display of the test information functions suggested that generally, the precision of measurement is maximized for middle ability distribution points. However, if item dependencies exist, then the accumulation of information around the middle of the distribution may be exaggerated. Further research should focus on understanding how information generated through the 3PL may be inflated due to local item dependencies.

Further, it would be useful to transform the current theta estimates into operational scores in order to obtain the classification accuracy provided by each model. Jiao, Liu, Haynie, Woo, and Gorham compared polytomous and dichotomous scoring algorithms for innovative items in a computer adaptive test (CAT) delivery context. They noted that classification rates were essentially the same between dichotomous (Rasch) and polytomous scoring methods for both

rater-generated and automated polytomous scoring algorithms under the PCM. As in this study, they found high correlations between estimated ability distributions for dichotomous and polytomous scores (.99). Given this, and the comparable ability distributions from the GPCM, GRM and operational 3PL, it is logical to expect similar, if not the same, classification rates. However, the topic should be empirically explored prior to making final conclusions about potential advantages to using a PIRT model to score examinees.

Goodness of fit tests are a necessary, but not sufficient criterion in the process of PIRT model selection. Samejima (1996) proposed additional criteria for evaluating models for polytomous responses. First and most paramount, she suggested that the psychological principle behind the model and its assumptions must match the cognitive situation under scrutiny. In the case of the CPA Exam TBSs, CPA candidates are exposed to several different types of psychological stimuli for which they receive different types of credit as evidenced by, for example, the existence of varying scoring rubrics for similar tasks. Thus, when selecting an appropriate mathematical model for such a psychological process, it is important to consider whether or not it becomes easier for a candidate of any ability to achieve the next summated category response as he or she accumulates more correct responses within a task (i.e., does the shape of the cumulative category response function change with increasing task score?). In the case of a "double-jeopardy" situation in which subsequent subtasks (MOs) depend to some extent upon successes on earlier sub-tasks/MOs, the shape of the conditional probability curve should vary across the summated task score range. Because TBSs vary in nature with regard to double jeopardy, it would be difficult to propose a single type of psychological process that dictates all responses to the CPA Exam simulation tasks.

In situations representing changing cumulative category response functions, a *step acceleration parameter* may be appropriate to model (Samejima, 1995). An acceleration model is appropriate for situations when the conditional probability curves change in shape across the ability continuum. In general, Samejima recommends the use of a "heterogeneous" model for response data that represents cognitive processes like problem solving, that is, a model that assumes heterogeneous step/threshold relationships with theta. An acceleration model as well as the GPCM are both examples of heterogeneous categorical models, whereas the GRM is an example of a homogenous model (Ostini & Nering, 2005; Samejima, 1996). Fit indices described in the current study should theoretically shed light on the question of whether a heterogeneous or homogenous model is more appropriate for modeling TBS data. The mixed results of this study, as indicated by Chen & Kang's $S\text{-}X^2$ fit index, behoove further exploration into the idea of heterogeneity. Future research should focus on modeling TBS response data using such a model to determine more precisely the extent to which characteristic curves are identical in shape (e.g., an assumption of the GRM).

The second criterion for model evaluation proposed by Samejima is *additivity*, which requires that combining two categories together results in the same operating mathematical model as the model for the original score; that is, item characteristic curves are identical in shape after some re-categorization of response options. In the situation of the Uniform CPA Exam TBSs, it would be interesting to give some attention to the visual comparison of the mathematical model before and after the summation of the dichotomous MOs. This is easily obtained using the Graphics package in PARSCALE 4.1 and the existing command files generated through this study. According to Samejima, Muraki's GPCM and Master's PCM are typical examples of heterogenous models, which become too complex to be able to satisfy the

additivity criterion, and should only be used in situations in which all response categories have particular absolute meanings (which is unlikely in the case of the Simulation tasks). An extension of the additivity criterion is the *natural generalization to a continuous response model* criterion. That is, as the number of response categories increases, the data may be manipulated using mathematical functions that apply to interval-type variables.

Finally, the last two proposed criteria for PIRT model evaluation are related to ability estimation. *Unique maximum* (likelihood) *condition* and *ordered modal points* are related to the idea that a person's ability may be defined by one and only one modal point of likelihood, and that per response category these modes are ordered in an ascending fashion. The latter two criteria are general to most polytomous models, but are still relevant in the study of the GRM and the GPCM. The item response information function can be used to ascertain the veracity of the last two criteria.

In practical terms, this study found that many of the TBS items do, indeed, fit either the GRM or the GPCM. However, there are several theoretical issues with these models. As explained by Samejima (1996) and Ostini and Nering (2005), homogenous models such as the GRM may not fit the practical reality of the complex cognitive processes at play during problem solving tasks, such as the TBSs. However, the strength of a homogenous model is that additivity always holds (Samjeima, 1996). On the other hand, the GPCM also has several potential issues, one of which is the changing meaning of the discrimination parameter depending on the number of response categories contained in a task. The changing meaning of the *a*-parameter may affect test assembly and the evaluation of the test information function because the discrimination parameter shrinks with an increasing number of response categories. Further, the GPCM does not satisfy Samejima's criteria for polytomous model evaluation. Specifically, the requirement of

additivity and generalizability to a continuous response model are not satisfied. Thus, if new polytomous response categories are added to any TBS, the cumulative category response function of the TBS may change in unpredictable ways, even if the new response categories are theoretically related to the measured latent construct. Further, as response options are added, the density characteristic for the continuous response will be invalid for the GPCM. However, being part of the heterogeneous case, both the PCM and GPCM offer modeling flexibility that may more accurately represent the complex psychological processes occurring during the Exam. Research into more flexible models such as the acceleration model may shed light on the more appropriate choice for modeling TBS responses.

References

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, *19*(6), 716-723.

Andrich, D. (1978a). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied psychological measurement*, *2*(4), 581-594.

Andrich, D. (1978b). A rating formulation for ordered response categories. *Psychometrika*, *43*(4), 561-573.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, *37*, 29-51.

Bond, T. G., & Fox, C. M. (2013). *Applying the Rasch model: Fundamental measurement in the human sciences*. Psychology Press.

Cai, L., Thissen, D., & du Toit, S. (2011). IRTPRO 2.1 for Windows. *Scientific Software International, Chicago, IL*.

De Ayala, R. J. (2009). *Theory and practice of item response theory*. Guilford Publications.

Demars, C. E. (2005). Type I error rates for PARSCALE's fit index. *Educational and psychological measurement*, *65*(1), 42-50.

Douglas, J., & Cohen, A. (2001). Nonparametric item response function estimation for assessing parametric model fit. *Applied Psychological Measurement*, *25*(3), 234-243.

Fitzpatrick, S.J., & Dodd, B.G. (1997). The effect on information of a transformation of the parameter scale. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Gelman A, Carlin J, Stern H, Rubin D (2003). *Bayesian Data Analysis*. CRC Press, Boca Raton, 2 edition.

Jansen, P. G., & Roskam, E. E. (1986). Latent trait models and dichotomization of graded responses. *Psychometrika, 51*(1), 69-91.

Jiao, H., Liu, J., Haynie, K., Woo, A., & Gorham, J. (2012). Comparison between dichotomous and polytomous scoring of innovative items in a large-scale computerized adaptive test. *Educational and Psychological Measurement*,*72*(3), 493-509.

Kang, T., & Chen, T. T. (2008). Performance of the Generalized S-X2 Item Fit Index for Polytomous IRT Models. *Journal of Educational Measurement, 45*(4), 391-406.

Kang, T., & Chen, T. T. (2011). Performance of the generalized *S-X2* item fit index for the graded response model. *Asia Pacific Education Review, 12*(1), 89-96.

Karabatsos, G. (1999). A critique of Rasch residual fit statistics. *Journal of Applied Measurement, 1*(2), 152-176.

Lee, G., Kolen, M. J., Frisbie, D. A., & Ankenmann, R. D. (2001). Comparison of dichotomous and polytomous item response models in equating scores from tests composed of testlets. *Applied Psychological Measurement, 25*(4), 357-372.

Li, S., & Wells, C. S. (2006, April). *A model fit statistic for Samejima's graded response model*. Paper presented at the annual meeting of National Council on Measurement in Education, San Francisco.

Liang, T. & Wells, C. S. (2009). A model fit statistic for generalized partial credit model. *Educational and Psychological Measurement, 69*(6), 913-928.

Linacre, J. M., & Wright, B. D. (1994). Chi-square fit statistics. *Rasch Measurement Transactions, 8*(2), 350.

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85-106.

Linacre, J. M. (2004). Estimation methods for Rasch measures. *Introduction to Rasch measurement*, 25-48.

Maydeu-Olivares, A., Drasgow, F., & Mead, A. D. (1994). Distinguishing among parametric item response models for polychotomous ordered data. Applied Psychological Measurement, 18(13), 245-56.

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

McKinley, R. L., & Mills, C. N. (1985). A comparison of several goodness-of-fit statistics. *Applied Psychological Measurement*, *9*, 49-57.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied psychological measurement*, *16*(2), 159-176.

Muraki, E. (1993). Information functions of the generalized partial credit model. *Applied Psychological Measurement*, *17*(4), 351-363.

Muraki, E., & Bock, R. D. (1997). PARSCALE: IRT item analysis and test scoring for rating scale data [Computer software]. Chicago: Scientific Software.

Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, *24*, 50-64.

Orlando, M., & Thissen, D. (2003). Further investigation of the performance of S-X2: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement*, *27*(4), 289-298.

Ostini, R., & Nering, M. L. (2005). *Polytomous item response theory models.* Thousand Oaks: Sage.

Penfield, R. D. (2014). An NCME Instructional Module on Polytomous Item Response Theory Models. *Educational Measurement: Issues and Practice*, *33*(1), 36-48.

Revuelta, J. (2005). An item response model for nominal data based on the rising selection ratios criterion. *Psychometrika*, *70*(2), 305-324.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applies statistician. *The Annals of Statistics*, *12*(4), 1151-1172.

Samejima, F. (1969). Estimation of latent ability groups using a response pattern of graded scores. *Psychometrika Monograph Supplement*, *4*, Part 2, Whole No. 17.

Samejima, F. (1972). A general model for free response data. *Psychometrika Monograph Supplement*, No. 18.

Samejima, F. (1996). Evaluation of mathematical models for ordered polychotomous responses. *Behaviormetrika, 23,* 17–35.

Scalise, K., Bernbaum, D. J., Timms, M., Harrell, S. V., Burmester, K., Kennedy, C. A., & Wilson, M. (2007). Adaptive technology for e-learning: principles and case studies of an emerging field. *Journal of the American Society for Information Science and Technology*, *58*(14), 2295-2309.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461-464.

Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the Reliability of Testlet-Based Tests. *Journal of Educational Measurement*, *28*(3), 237-247.

Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology*, *8*(1), 33.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B, 64*, 583–639.

Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). WinBUGS user manual. *Cambridge: MRC Biostatistics Unit*.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied psychological measurement*, *7*(2), 201-210.

Stone, C. A., & Hansen, M. A. (2000). The effect of errors in estimating ability on goodness-of-fit tests for IRT models. *Educational and Psychological Measurement*, *60*(6), 974-991.

Stone, C. A., & Hansen, M. A. (2000). The effect of errors in estimating ability on goodness-of fit tests for IRT models. *Educational and Psychological Measurement*, *60*, 974-991.

Thissen, D., Pommerich, M., Billeaud, K., & Williams, V. S. (1995). Item response theory for scores on tests including polytomous items with ordered responses. *Applied Psychological Measurement*, *19*(1), 39-49.

Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika*, *51*(4), 567-577.

Thissen, D., Steinberg, L., & Mooney, J. A. (1989). Trace Lines for Testlets: A Use of Multiple-Categorical-Response Models. *Journal of Educational Measurement*, *26*(3), 247-260.

Thissen, D., Steinberg, L., & Fitzpatrick, A. R. (1989). Multiple-Choice Models: The Distractors Are Also Part of the Item. *Journal of Educational Measurement*,*26*(2), 161-176.

Wainer, H. (1995). Precision and differential item functioning on a testlet-based test: The 1991 Law School Admissions Test as an example. *Applied Measurement in Education*, *8*(2), 157-86.

Wainer, H., & Kiely, G. L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational measurement*, *24*(3), 185-201.

Wainer, H., & Lewis, C. (1990). Toward a psychometrics for testlets. *Journal of Educational Measurement*, *27*(1), 1-14.

Wainer, H., & Thissen, D. (1996). How is reliability related to the quality of test scores? What is the effect of local dependence on reliability?. *Educational Measurement: Issues and Practice*, *15*(1), 22-29.

Wang, W. C., & Chen, C. T. (2005). Item parameter recovery, standard error estimates, and fit statistics of the WINSTEPS program for the family of Rasch models. *Educational and Psychological Measurement*, *65*(3), 376-404.

Wright, B. D., & Panchapakesan, N. (1969). A procedure for sample-free item analysis. *Educational and Psychological Measurement*, *29*, 23-48.

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

Wright, B. D.,& Mead, R. J. (1977). *BICAL: Calibrating items and scales with the Rasch model* (Research memorandum No. 23). Chicago: University of Chicago, Department of Education, Statistical Laboratory.

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

Yen, W. M. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement*, *5*, 245-262.

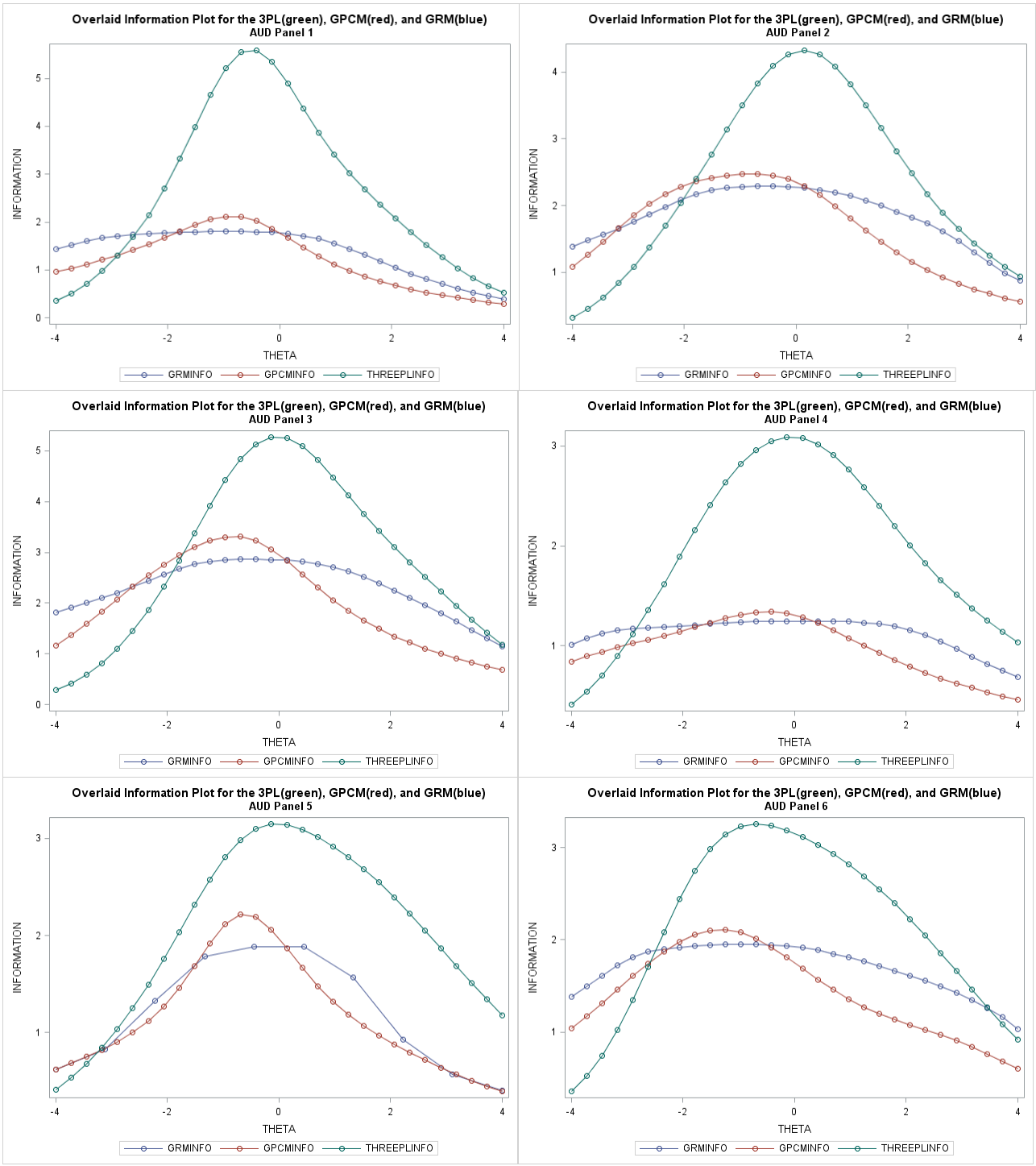
Yurekli, H. (2010). The relationship between parameters from some polytomous item response theory models (Master's thesis). Retrieved from http://diginole.lib.fsu.edu/etd/1104

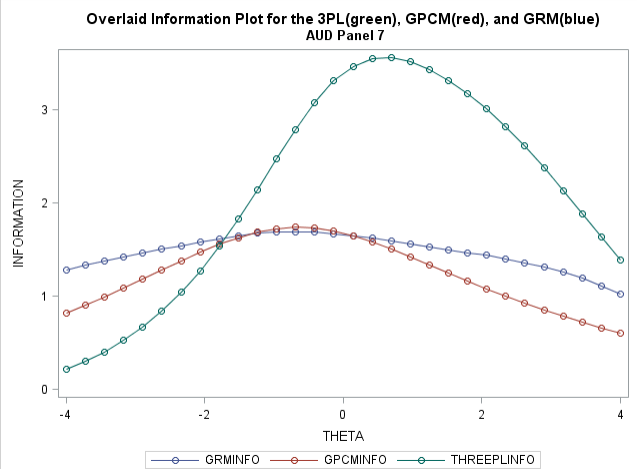
Zhu, X., & Stone, C. A. (2011). Assessing fit of unidimensional graded response models using Bayesian methods. *Journal of Educational Measurement*, *48*(1), 81-97.

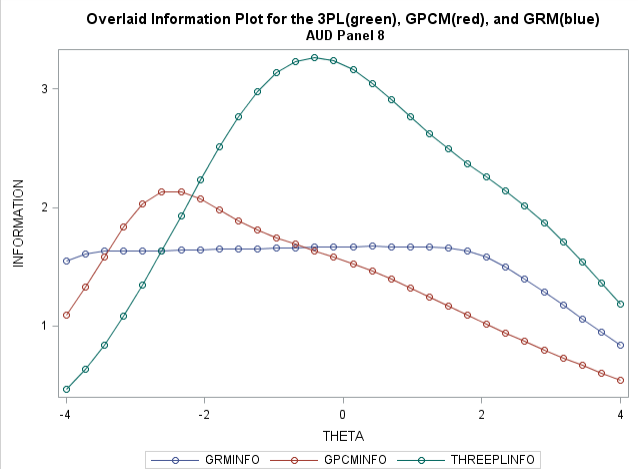
Appendix A

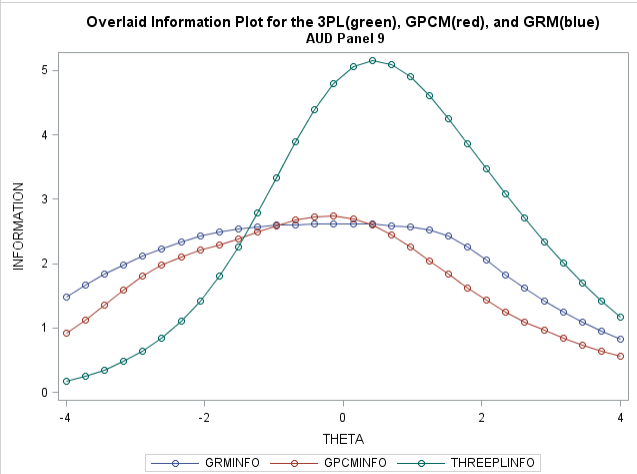
Information Functions of Panel Analyzed Using the GRM, GPCM, and the 3PL

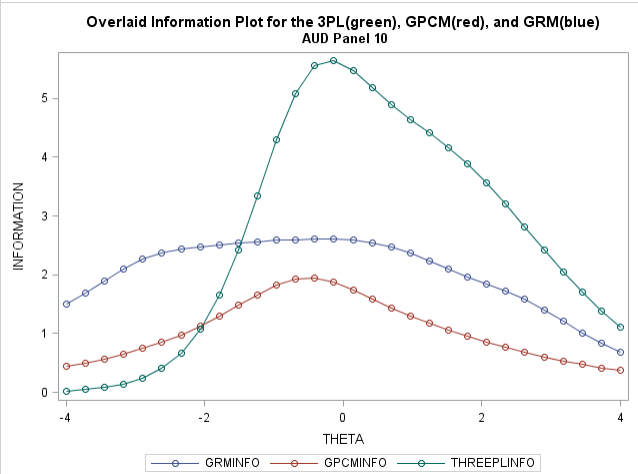
AUD Section

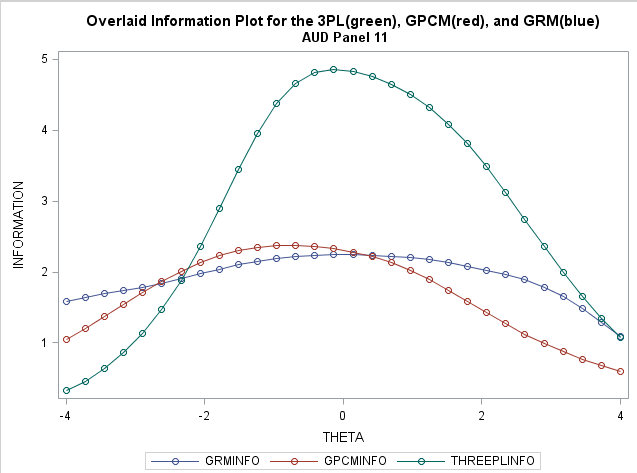


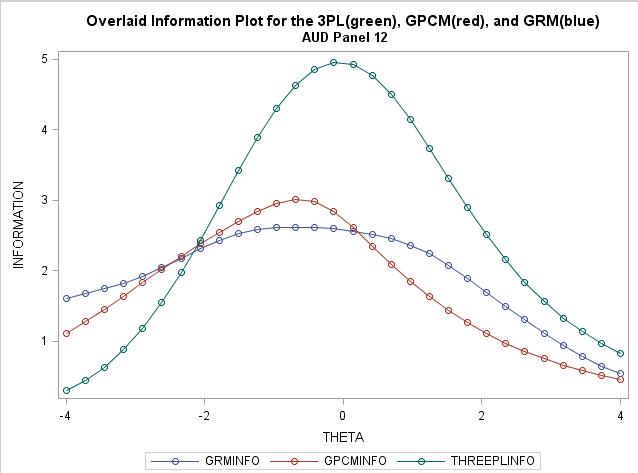

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
AUD Panel 7
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

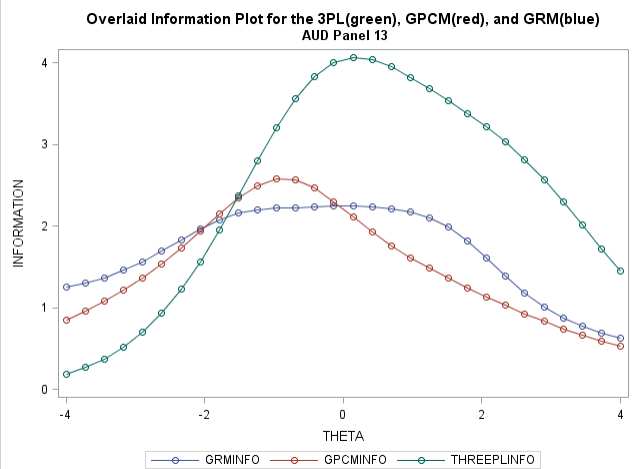

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
AUD Panel 8
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

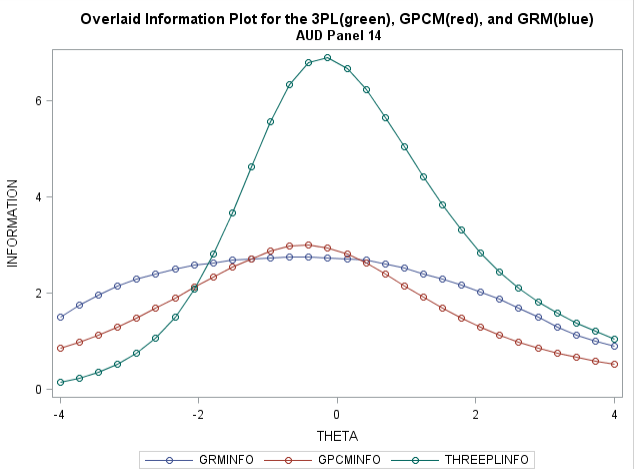

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
AUD Panel 9
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

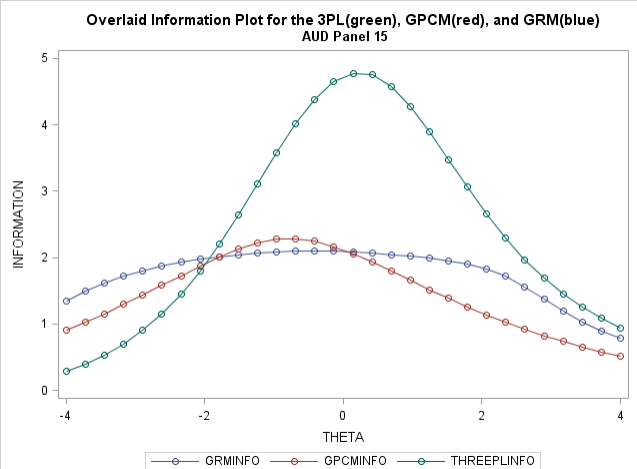

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
AUD Panel 10
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

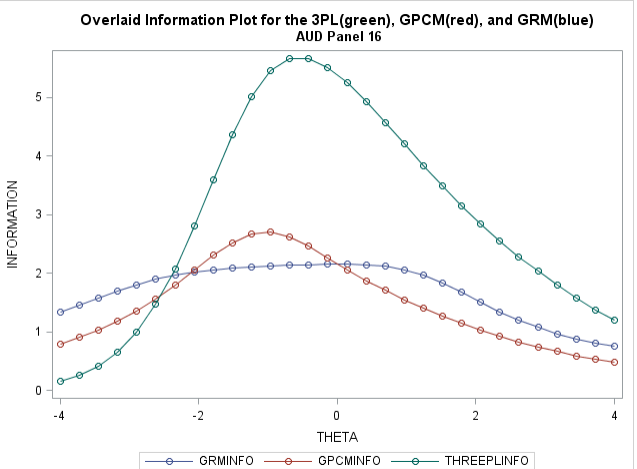

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
AUD Panel 11
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

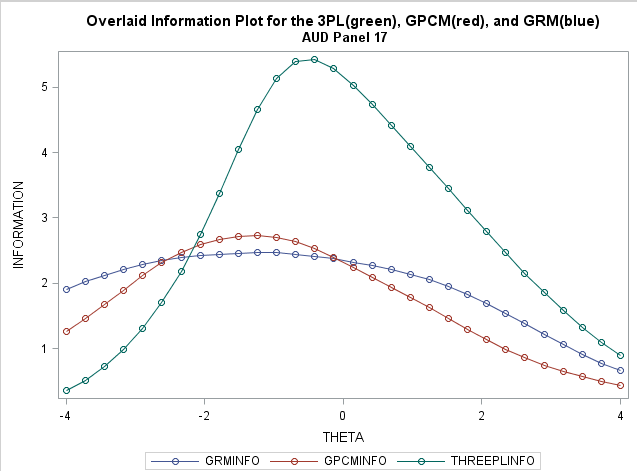

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
AUD Panel 12
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

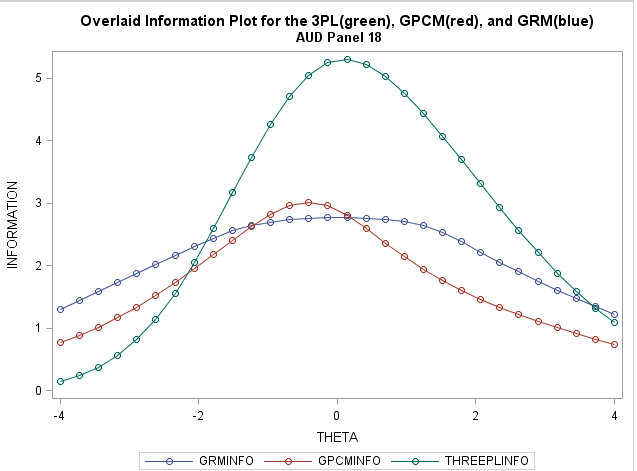

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
AUD Panel 13
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

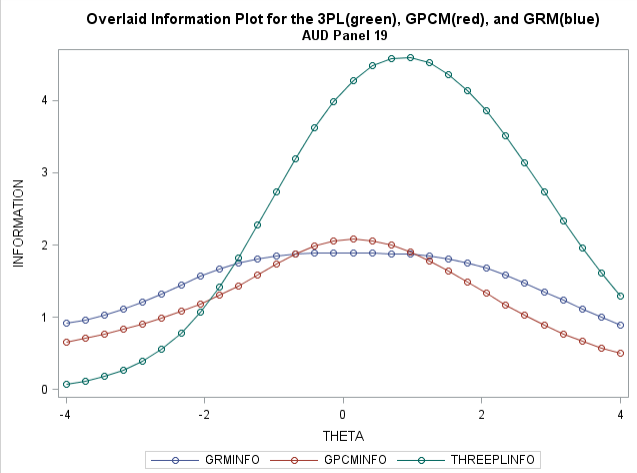

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
AUD Panel 14
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

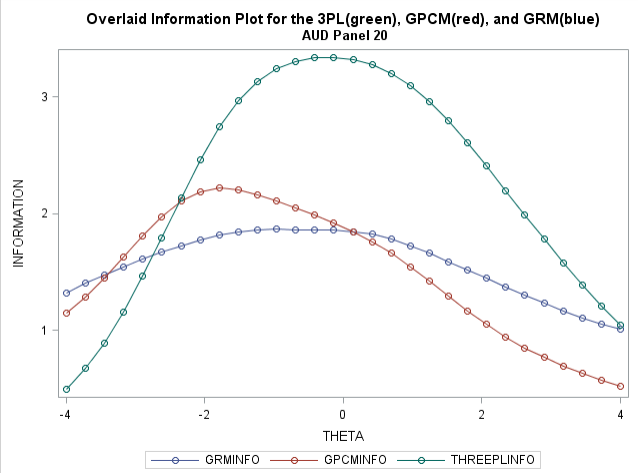

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
AUD Panel 15
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

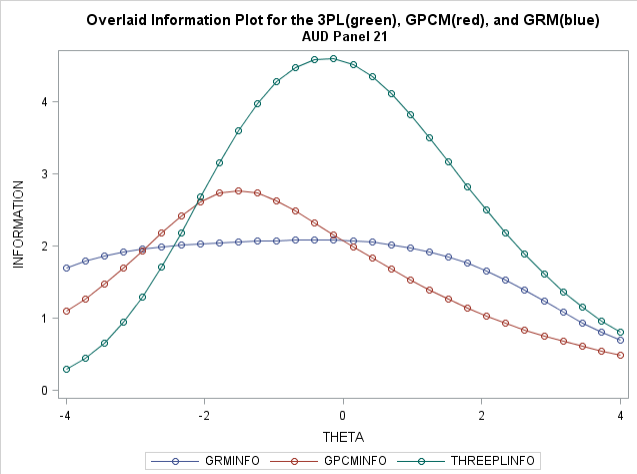

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
AUD Panel 16
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

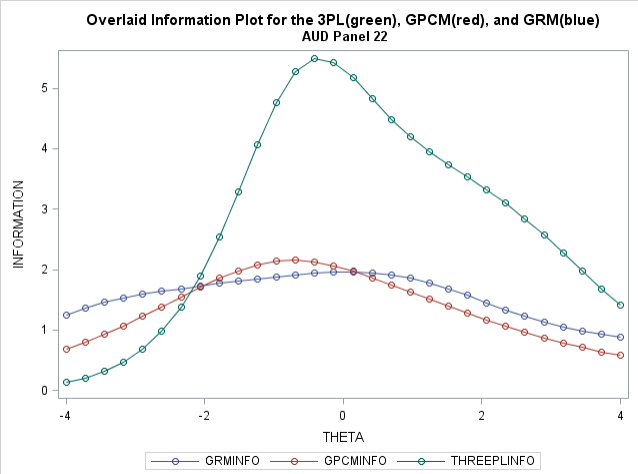

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
AUD Panel 17
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

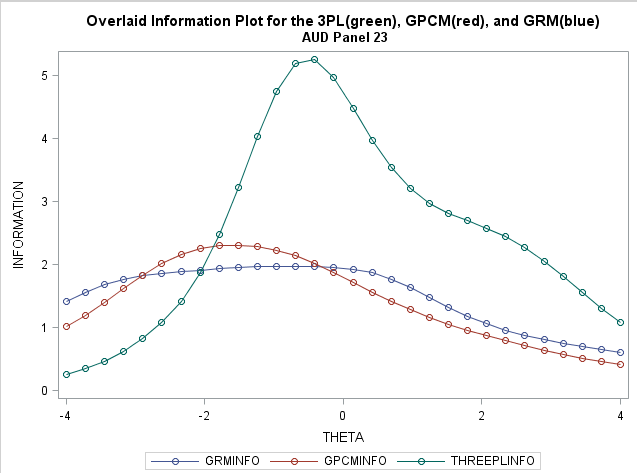

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
AUD Panel 18
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

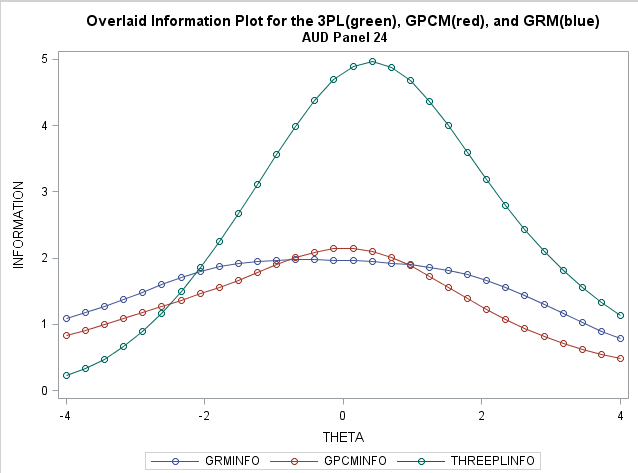

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
AUD Panel 19
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO


Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
AUD Panel 20
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

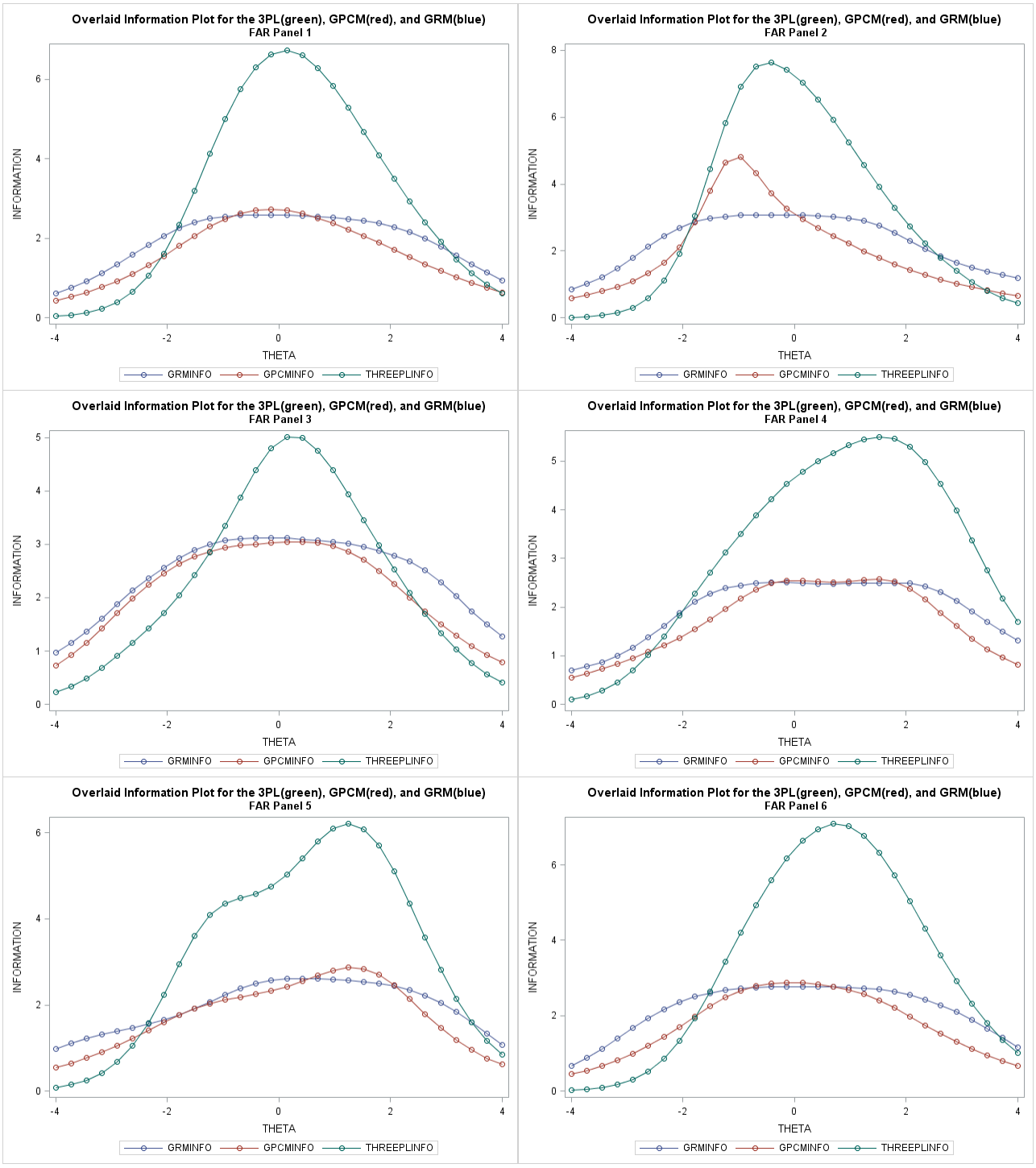

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
AUD Panel 21
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

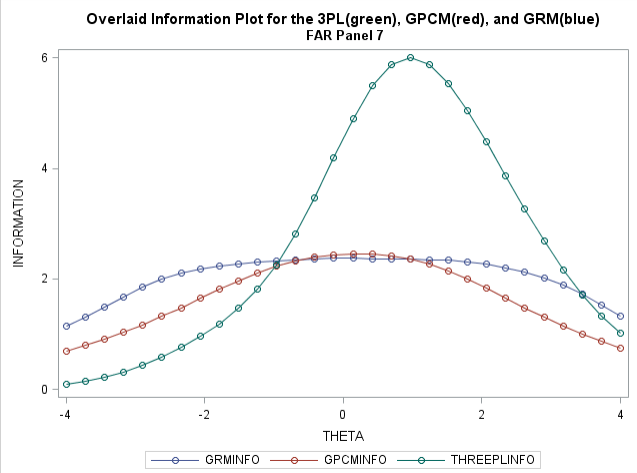

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
AUD Panel 22
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

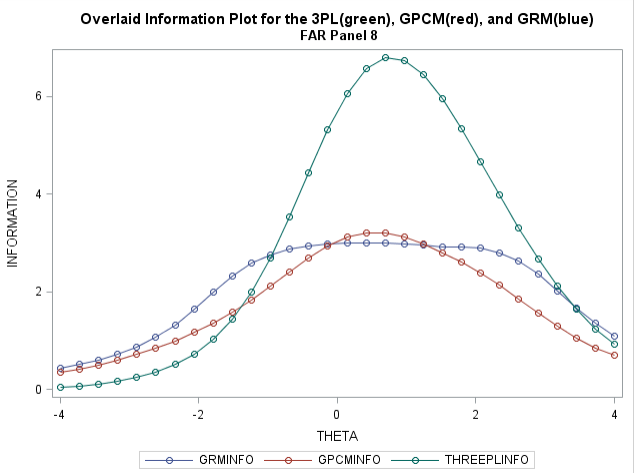

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
AUD Panel 23
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

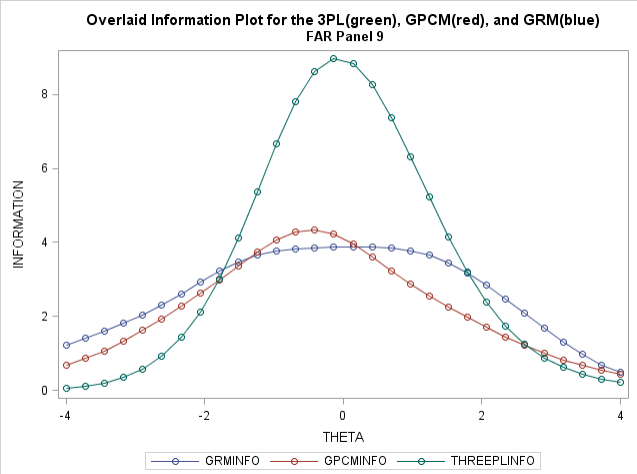

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
AUD Panel 24
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

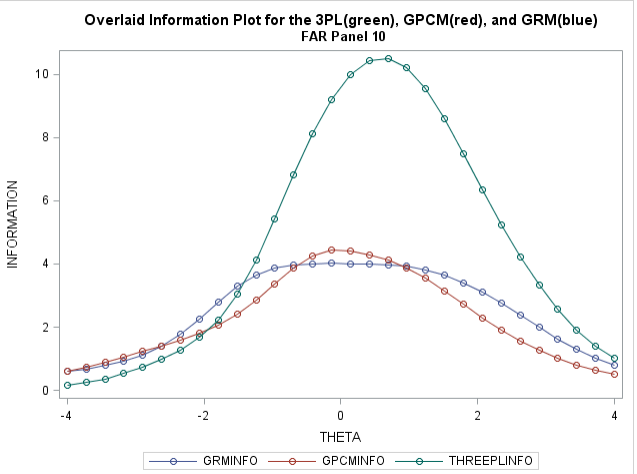
Information Functions of Panel Analyzed Using the GRM, GPCM, and the 3PL
FAR Section

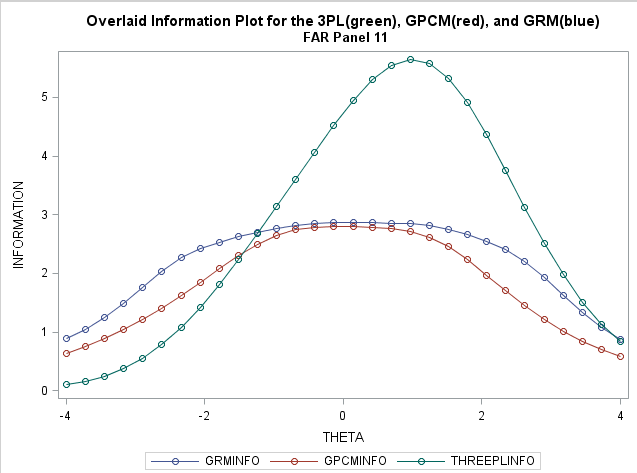


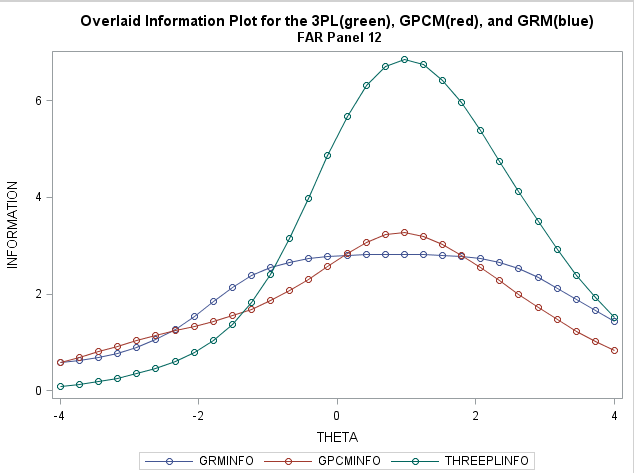

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
FAR Panel 7
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

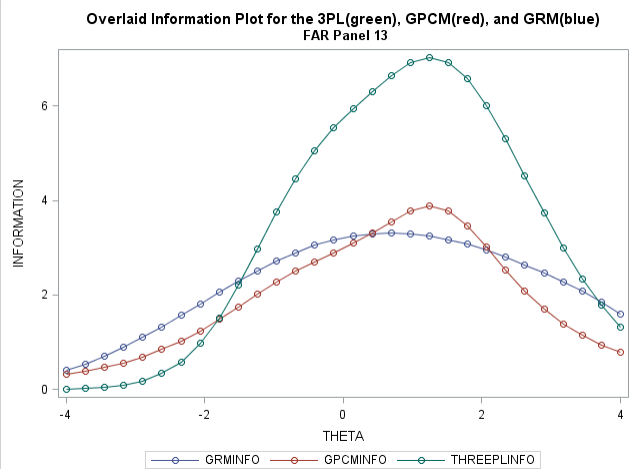

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
FAR Panel 8
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

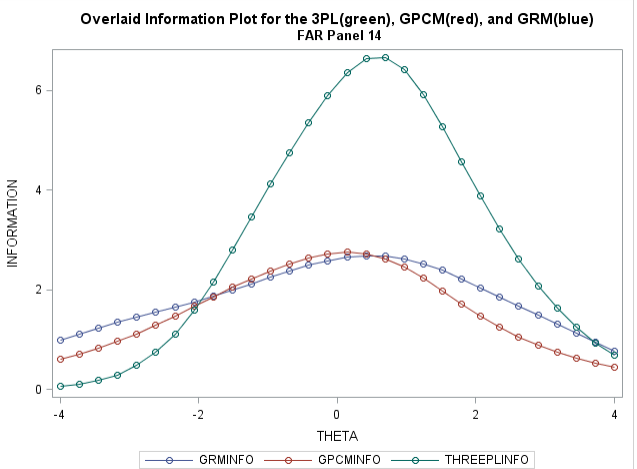

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
FAR Panel 9
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

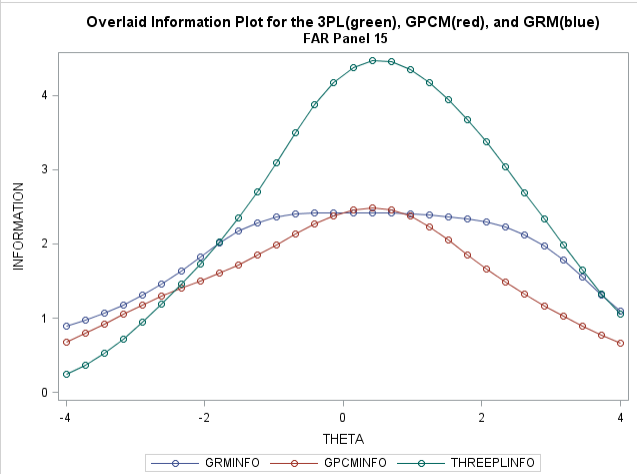

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
FAR Panel 10
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

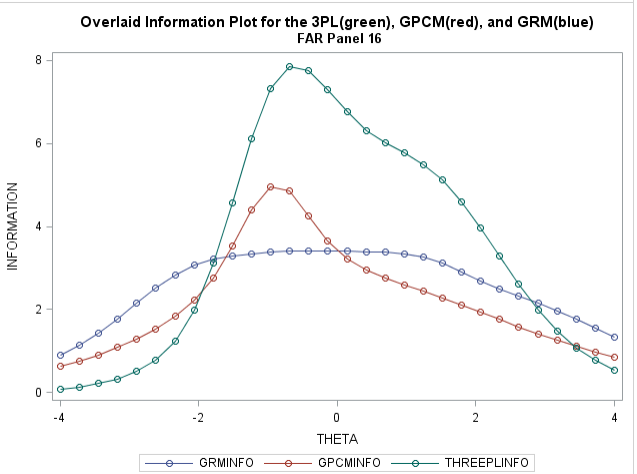

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
FAR Panel 11
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

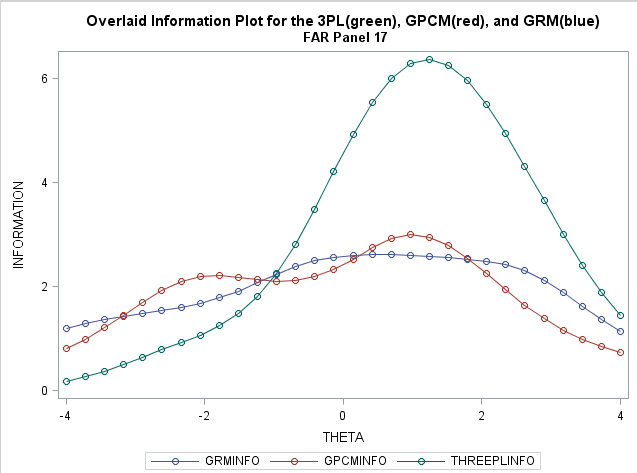

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
FAR Panel 12
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

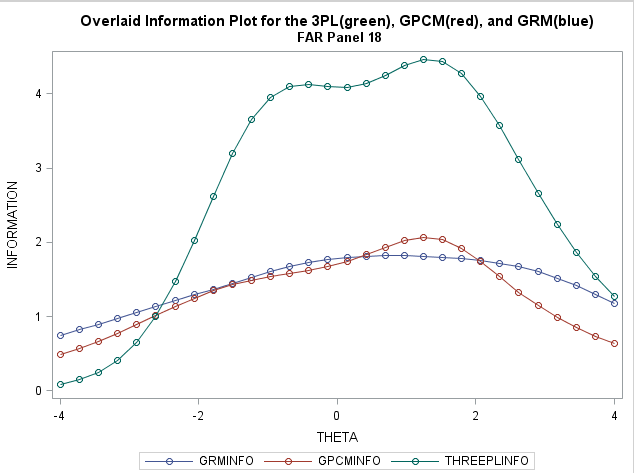

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
FAR Panel 13
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

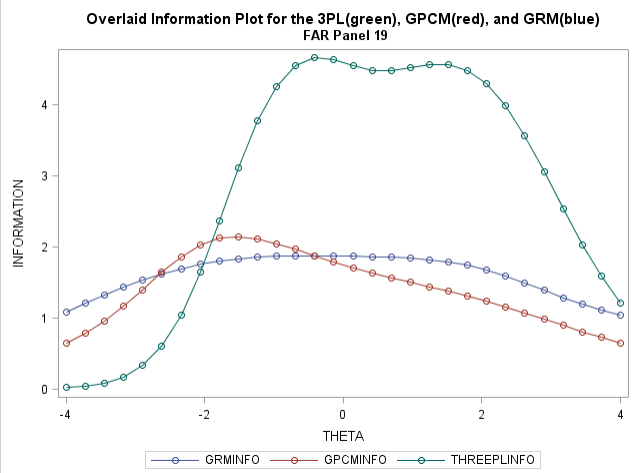

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
FAR Panel 14
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

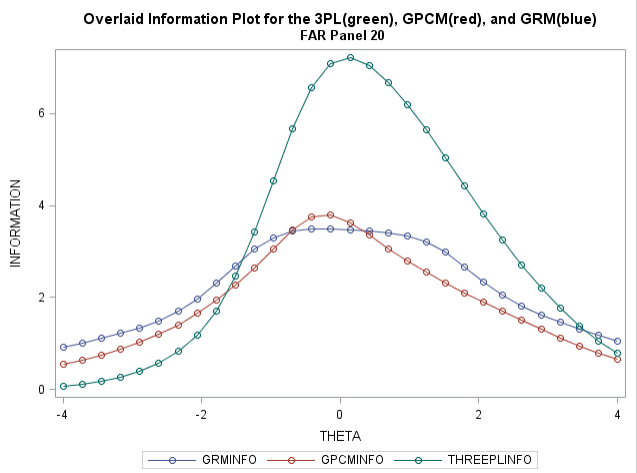

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
FAR Panel 15
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

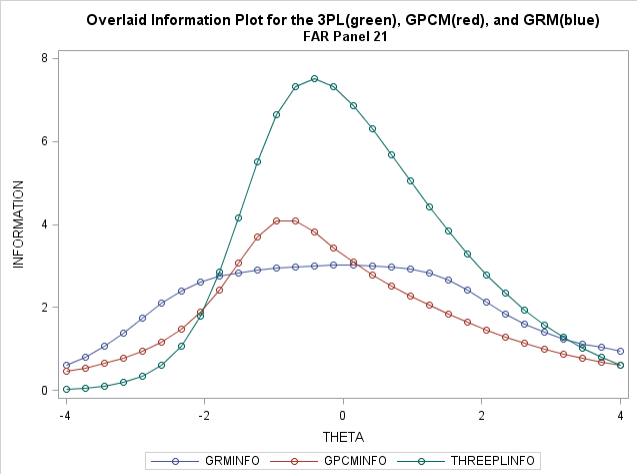

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
FAR Panel 16
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

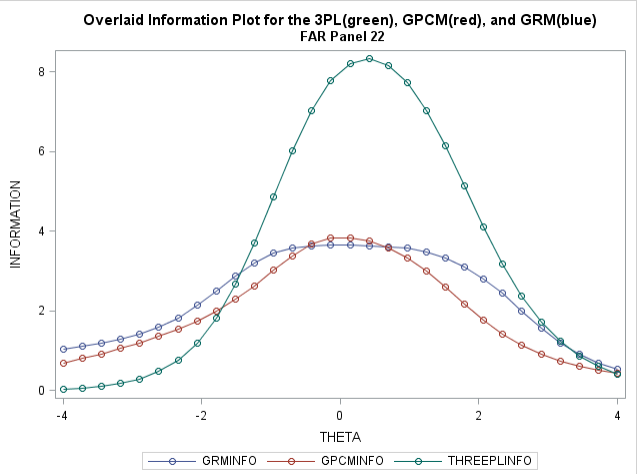

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
FAR Panel 17
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

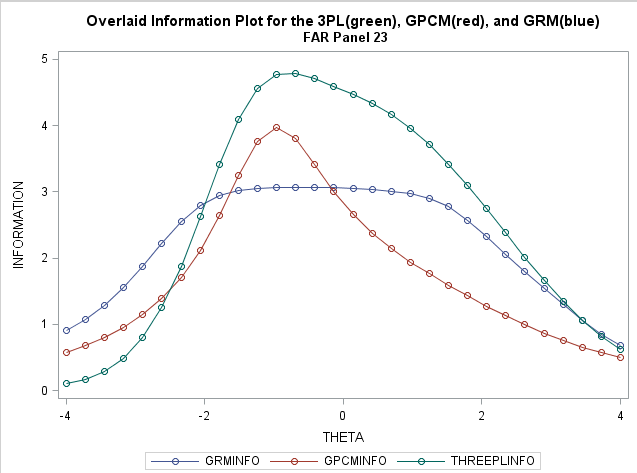

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
FAR Panel 18
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

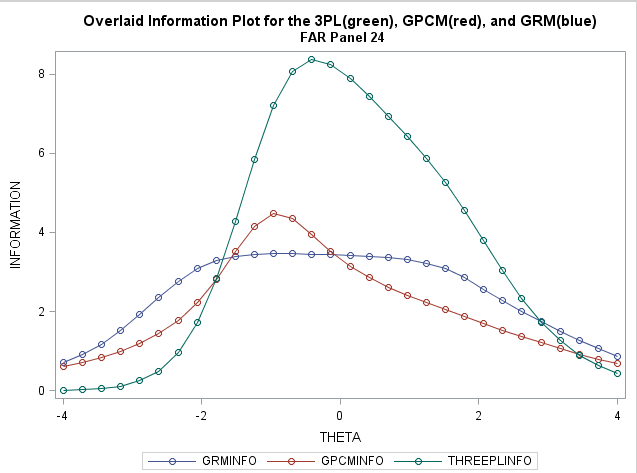

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
FAR Panel 19
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO


Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
FAR Panel 20
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

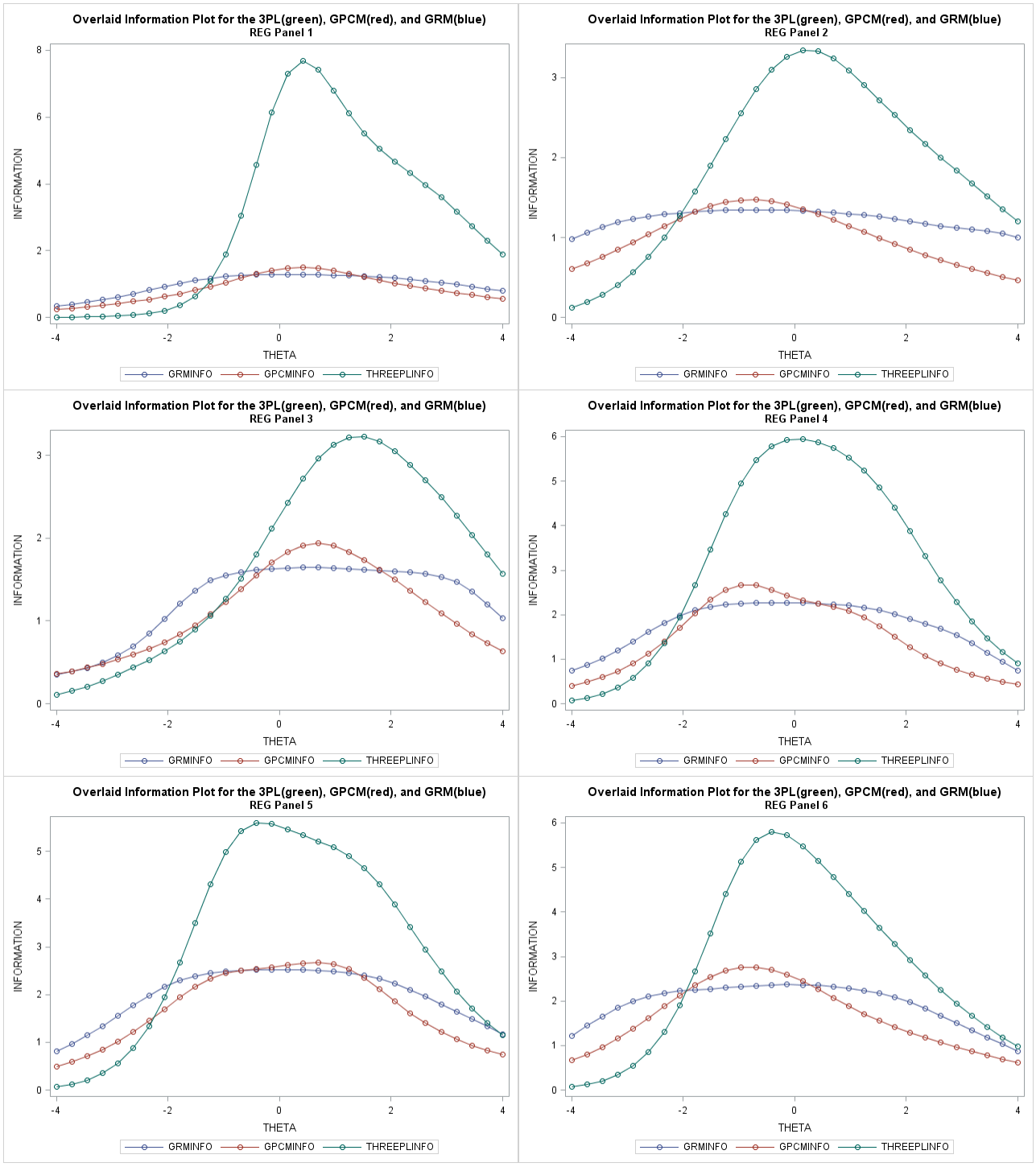

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
FAR Panel 21
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

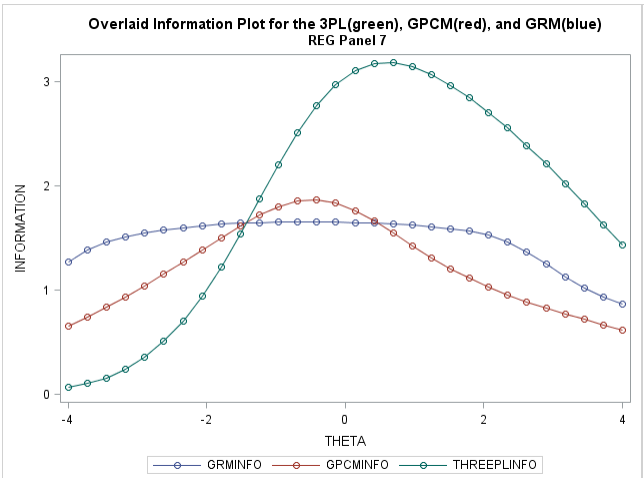

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
FAR Panel 22
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

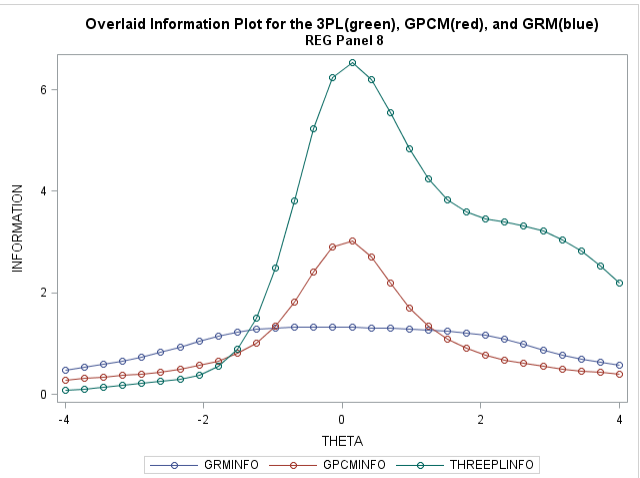

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
FAR Panel 23
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

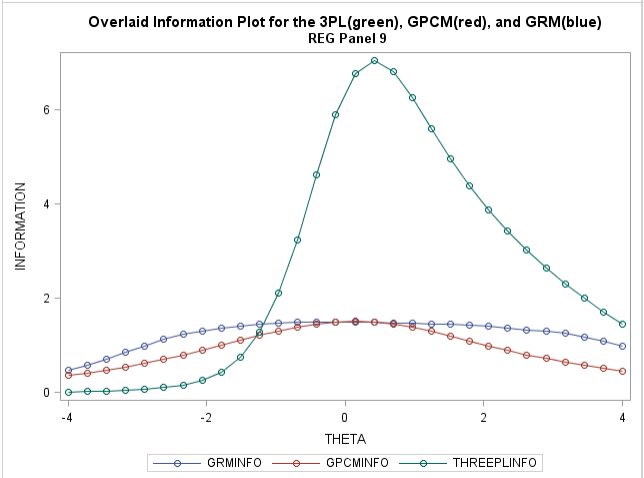

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
FAR Panel 24
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

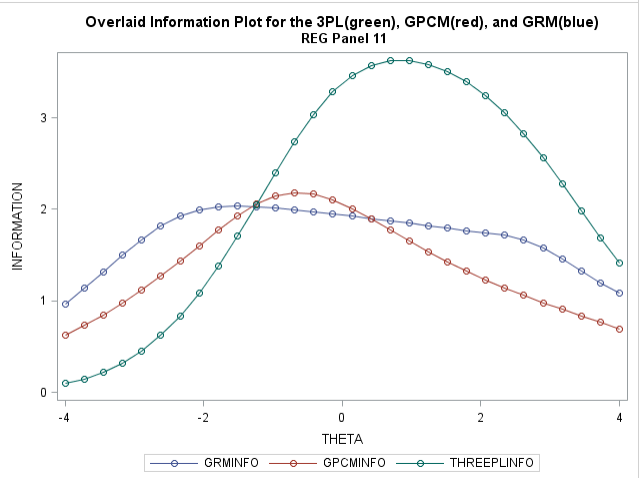
Information Functions of Panel Analyzed Using the GRM, GPCM, and the 3PL
REG Section

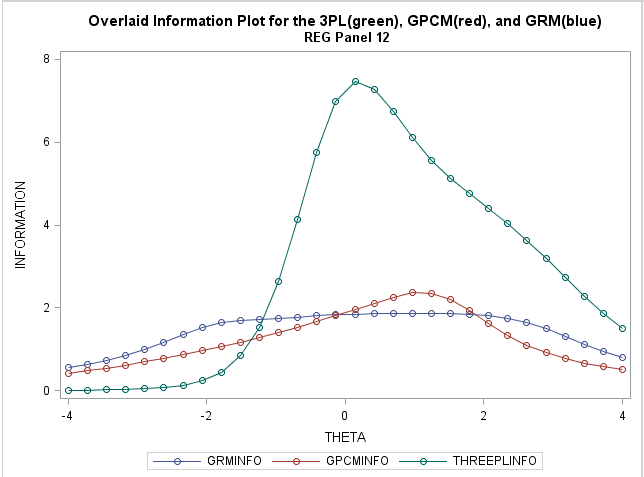


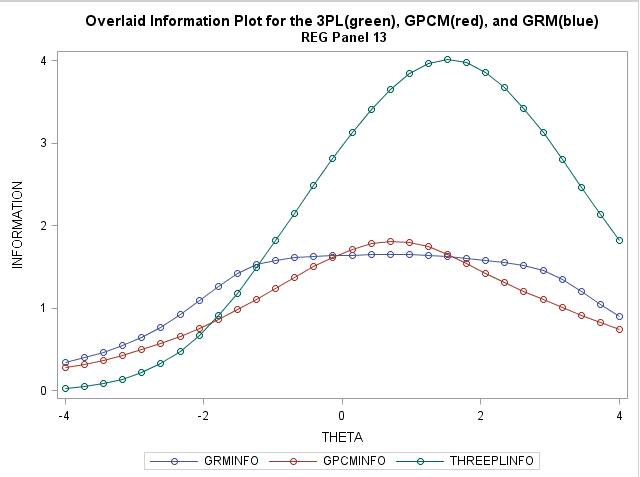

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
REG Panel 7
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

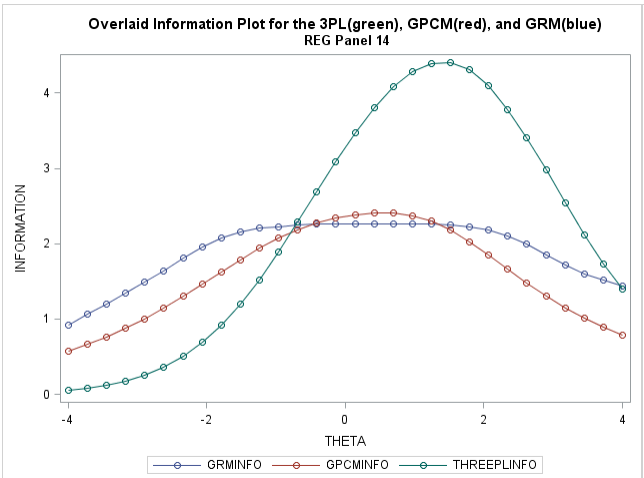

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
REG Panel 8
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

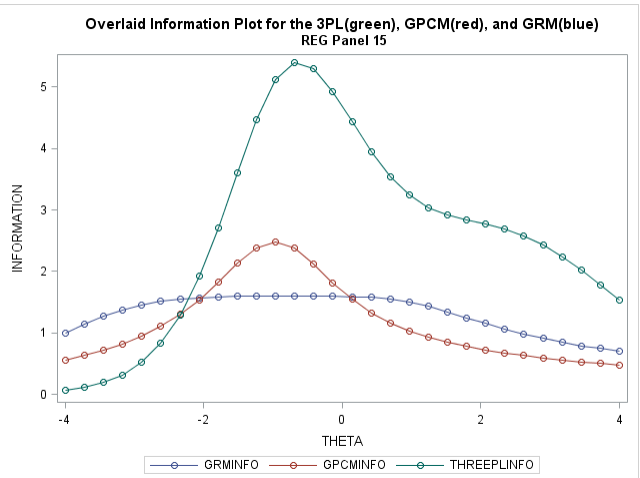

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
REG Panel 9
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

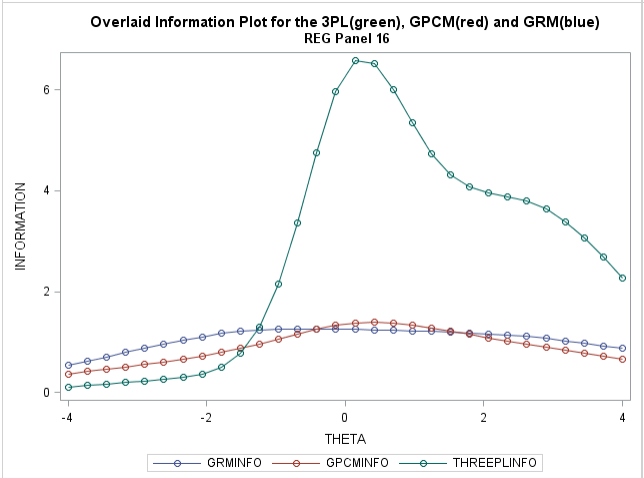

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
REG Panel 11
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

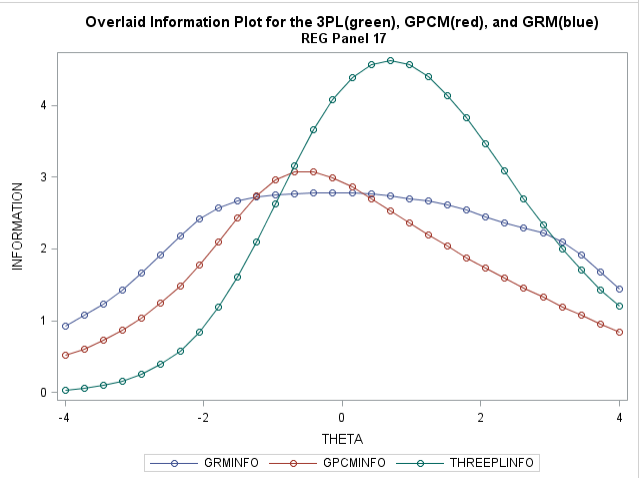

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
REG Panel 12
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

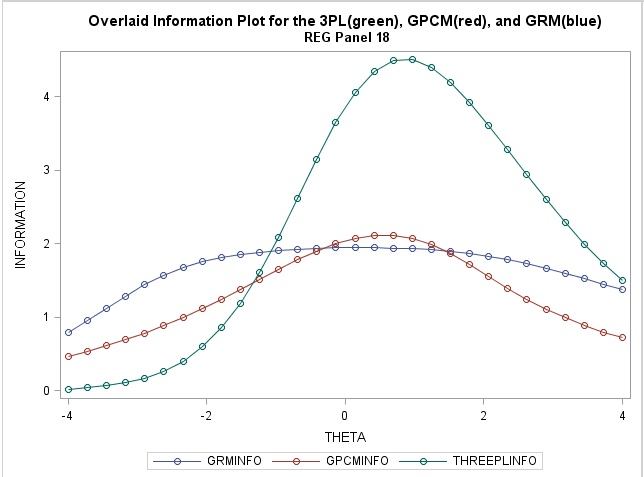

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
REG Panel 13
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

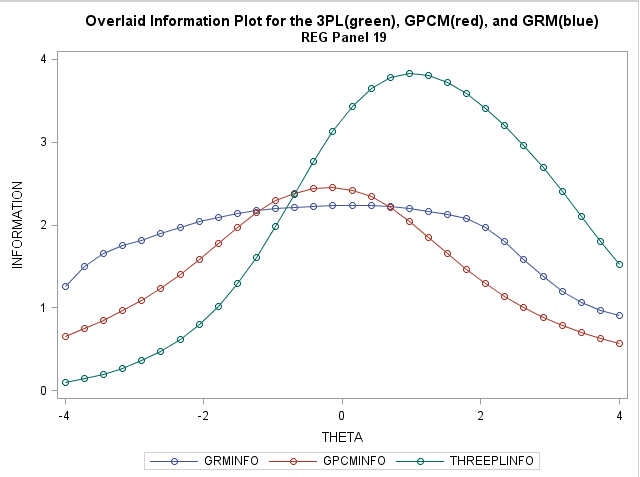

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
REG Panel 14
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

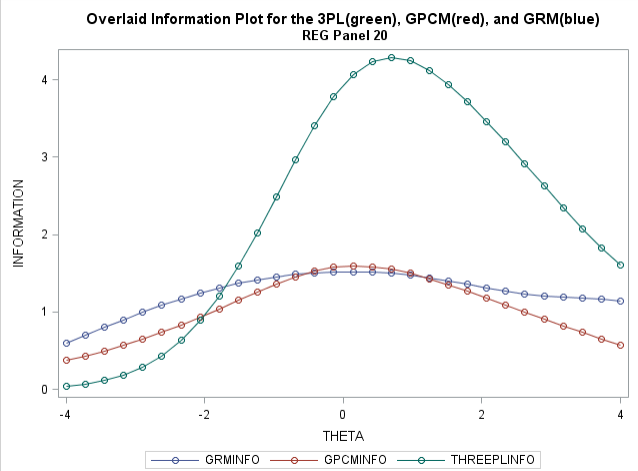

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
REG Panel 15
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

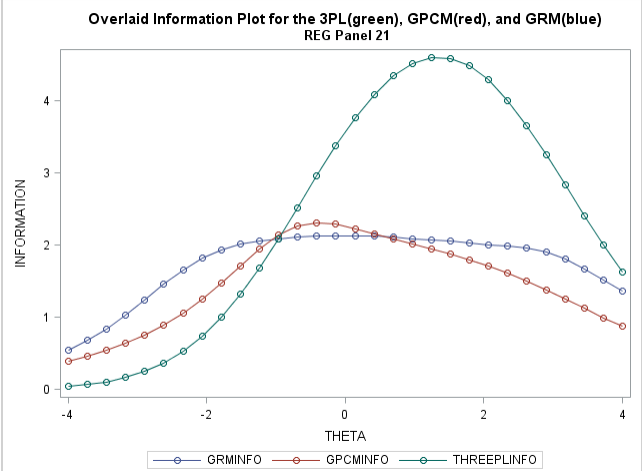

Overlaid Information Plot for the 3PL(green), GPCM(red) and GRM(blue)
REG Panel 16
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

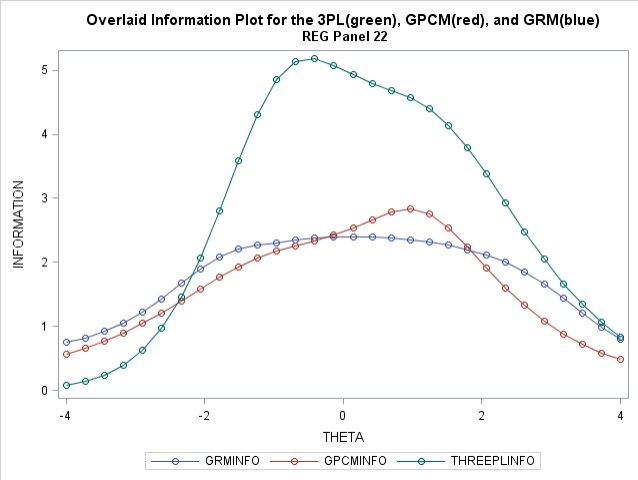

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
REG Panel 17
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

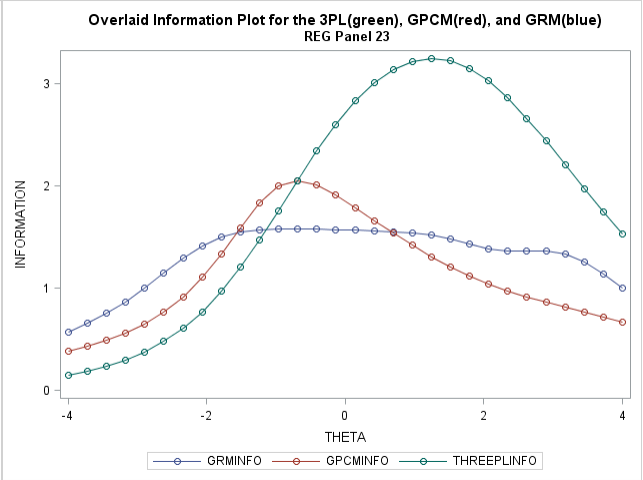

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
REG Panel 18
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO

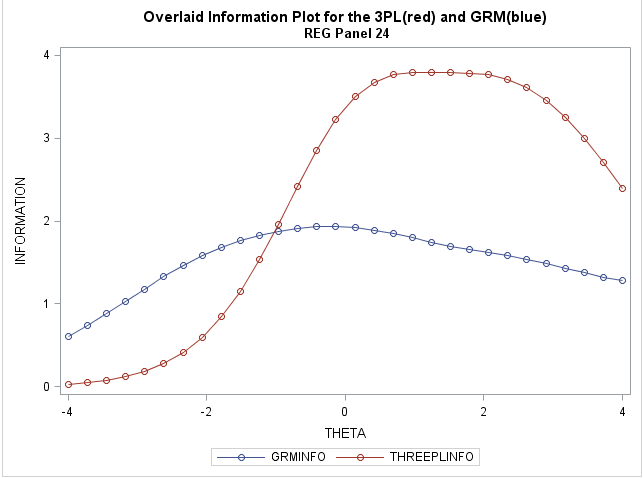

Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
REG Panel 19
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO


Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
REG Panel 20
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO


Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
REG Panel 21
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO


Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
REG Panel 22
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO


Overlaid Information Plot for the 3PL(green), GPCM(red), and GRM(blue)
REG Panel 23
INFORMATION
THETA
GRMINFO
GPCMINFO
THREEPLINFO


Overlaid Information Plot for the 3PL(red) and GRM(blue)
REG Panel 24
INFORMATION
THETA
GRMINFO
THREEPLINFO

Appendix B

S-X^2 Fit Statistics for Panel Data

	GRM			1PLGRM			GPCM			PCM		
Label	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability
REG1	118.5	117	0.445	122.42	121	0.448	115.42	116	0.498	117.88	116	0.435
REG1	84.97	92	0.686	102.57	100	0.410	87.27	93	0.649	86.16	94	0.706
REG1	95.32	98	0.558	102.86	103	0.486	90.68	96	0.635	91.29	99	0.697
REG1	97.7	102	0.603	95.25	102	0.669	96.99	100	0.567	98.16	100	0.534
REG1	93.35	79	0.129	116.28	86	0.017	96.85	79	0.084	101.34	86	0.124
REG10	81.88	85	0.576	82.12	86	0.599	78.75	85	0.671	82.34	83	0.501
REG10	95.62	98	0.550	101.43	96	0.332	102.13	97	0.341	102.58	96	0.304
REG10	79.5	78	0.433	77.97	76	0.417	74.34	76	0.533	79	74	0.323
REG10	86.27	83	0.381	87.8	82	0.310	86.82	84	0.394	88.73	79	0.212
REG11	157.5	139	0.135	166.81	144	0.094	147.55	136	0.235	145.96	134	0.226
REG11	166.26	121	0.004	169.94	125	0.005	155.68	118	0.012	164.98	111	0.001
REG11	147.4	133	0.186	148.08	134	0.191	124.68	129	0.592	139.95	121	0.115
REG11	77.28	97	0.930	86.95	104	0.886	70.03	93	0.964	71.77	93	0.950
REG11	102.8	112	0.722	101.82	113	0.766	87.07	113	0.967	116.33	104	0.192
REG12	121.69	101	0.079	129.85	107	0.066	125.71	101	0.048	132.4	105	0.036
REG12	97.66	94	0.377	103.32	95	0.263	100.23	90	0.216	98.96	91	0.266
REG12	95.98	88	0.263	114.3	99	0.139	95.24	90	0.332	117.19	96	0.070
REG12	75.3	82	0.687	75.59	82	0.679	83.27	80	0.379	82.4	82	0.468
REG12	80.49	85	0.619	89.88	91	0.514	80.93	80	0.451	88.76	84	0.340
REG13	103.08	94	0.245	108.21	96	0.186	99.39	91	0.257	109.48	94	0.131
REG13	79.02	76	0.383	82.74	77	0.306	78.99	75	0.354	84.87	76	0.227
REG13	73.1	81	0.723	67.51	81	0.858	72.17	81	0.748	70.41	81	0.794
REG13	71.06	73	0.543	72.44	75	0.563	76.42	72	0.338	75.88	73	0.385
REG14	90.81	82	0.237	92.1	83	0.231	78.49	81	0.559	82.94	82	0.451
REG14	63.5	65	0.530	64.99	65	0.478	61.22	62	0.505	61.89	62	0.481
REG14	87.45	66	0.040	91.97	67	0.023	80.78	64	0.077	83.39	64	0.052
REG14	73.99	73	0.447	82.31	70	0.149	65.33	70	0.636	81.46	67	0.110
REG14	63.7	63	0.453	80.83	65	0.089	59.15	63	0.615	63.62	61	0.384
REG15	94.24	93	0.445	103.1	98	0.342	87.96	92	0.600	99.23	97	0.418
REG15	66.66	76	0.770	72.67	78	0.650	67.94	74	0.677	69.01	75	0.673
REG15	77.72	90	0.819	92.6	92	0.464	83.82	90	0.664	100.99	95	0.317
REG15	81.37	82	0.500	90.38	86	0.352	79.89	80	0.483	80.01	83	0.573
REG15	76.81	72	0.327	82.73	74	0.228	77.18	73	0.346	77.21	73	0.345
REG16	94.73	86	0.243	94.29	85	0.230	93.01	84	0.235	95.63	82	0.144
REG16	91.32	88	0.383	91.25	91	0.474	92.22	89	0.386	96.16	87	0.235
REG16	63.21	92	0.991	60.92	94	0.997	58.49	90	0.996	57.67	90	0.997
REG16	85.31	79	0.294	84.45	79	0.316	87.44	78	0.217	88.91	77	0.166
REG17	57.02	57	0.475	83.44	61	0.030	54.4	54	0.460	69.71	60	0.183

	GRM			1PLGRM			GPCM			PCM		
Label	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability
REG1	118.5	117	0.445	122.42	121	0.448	115.42	116	0.498	117.88	116	0.435
REG17	90.57	90	0.464	100.94	92	0.246	87.91	86	0.423	94.45	89	0.326
REG17	73.7	73	0.456	96.69	79	0.086	76.1	73	0.379	83.11	76	0.270
REG17	48.39	75	0.993	61.39	83	0.964	52.39	73	0.967	62.51	82	0.946
REG17	76.93	84	0.695	82.83	87	0.607	67.91	81	0.850	69.5	82	0.836
REG18	11.67	19	0.900	14.65	19	0.745	11.67	19	0.900	32.08	20	0.042
REG18	12.07	20	0.914	19.8	20	0.472	12.4	20	0.902	12.8	20	0.886
REG18	14.04	20	0.829	29.05	18	0.048	14.12	20	0.825	15.25	20	0.763
REG19	30.26	23	0.142	43.86	21	0.002	30.69	23	0.130	37.62	24	0.038
REG19	7.88	21	0.996	18.47	19	0.493	8.2	21	0.994	13.11	22	0.930
REG19	24.97	18	0.125	28.99	18	0.048	24.8	18	0.130	31.63	18	0.024
REG19	17.32	21	0.692	30.1	21	0.090	17.62	21	0.674	20.94	22	0.526
REG2	73.58	84	0.785	72.32	84	0.815	78.08	83	0.633	81.39	83	0.530
REG2	113.5	88	0.035	113.29	88	0.036	114.57	86	0.021	114.76	86	0.021
REG2	92.22	84	0.252	92.15	85	0.279	99.24	87	0.174	94.53	86	0.248
REG2	95.43	79	0.100	103.01	85	0.089	98.93	82	0.098	94.61	81	0.143
REG20	11.48	13	0.572	12.51	15	0.641	11.68	13	0.555	17.42	15	0.294
REG20	18.07	15	0.258	20.66	16	0.192	17.88	15	0.268	30.25	18	0.035
REG20	21.67	16	0.154	22.32	16	0.133	22.74	16	0.121	29	17	0.035
REG21	15.78	19	0.673	16.2	19	0.645	15.71	19	0.677	29.24	21	0.108
REG21	9.61	22	0.990	20.89	22	0.529	10.21	22	0.984	13.28	23	0.946
REG21	18.09	17	0.385	18.01	17	0.390	18.22	17	0.377	37.78	18	0.004
REG22	11.18	15	0.740	19.69	16	0.234	11.18	15	0.740	11.11	16	0.803
REG22	41.87	18	0.001	40.72	19	0.003	66.05	17	0.000	21.35	20	0.378
REG22	21.12	21	0.453	56.61	20	0.000	21.15	21	0.451	21.29	21	0.443
REG23	21.06	17	0.223	24.78	18	0.131	21.06	17	0.223	25.94	18	0.101
REG23	24.75	26	0.534	88.13	25	0.000	24.75	26	0.534	26.07	27	0.516
REG23	30.65	19	0.044	48.97	18	0.000	30.79	19	0.042	30.78	20	0.058
REG23	21.3	22	0.504	52.65	21	0.000	21.88	24	0.588	22.86	24	0.530
REG23	24.52	18	0.138	29.56	17	0.030	24.78	18	0.131	25.21	18	0.119
REG24	71.31	54	0.057	74.78	52	0.021	66.57	52	0.084	78.88	50	0.006
REG24	79.32	64	0.094	88.77	61	0.012	77.1	63	0.109	81.88	62	0.046
REG24	76.23	61	0.090	86.02	60	0.015	71.74	61	0.163	76.95	60	0.069
REG24	73.34	69	0.337	103.28	65	0.002	70.82	69	0.418	86.4	66	0.047
REG24	70.17	66	0.339	78.4	63	0.091	62.7	64	0.524	73.08	64	0.204
REG25	36.25	22	0.029	103.52	21	0.000	36.21	22	0.029	39.75	22	0.012
REG25	20.07	20	0.455	25.67	20	0.176	16.26	20	0.701	18.41	21	0.624
REG26	92.59	82	0.199	90.42	87	0.379	88.93	82	0.281	79.55	83	0.588
REG26	94.47	93	0.439	93.89	93	0.455	84.93	91	0.660	87.81	91	0.576
REG26	102.21	105	0.559	101.46	106	0.607	99.46	104	0.608	97.47	105	0.687

	GRM			1PLGRM			GPCM			PCM		
Label	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability
REG1	118.5	117	0.445	122.42	121	0.448	115.42	116	0.498	117.88	116	0.435
REG26	91.9	89	0.395	93.72	91	0.401	94.54	91	0.379	96.04	91	0.338
REG27	141.51	120	0.088	153.44	129	0.070	123.42	117	0.324	127.63	123	0.369
REG27	78.88	84	0.638	86.63	88	0.522	78.35	83	0.624	77.77	85	0.699
REG27	88.26	84	0.354	97.49	91	0.302	85.46	83	0.404	85.6	85	0.462
REG27	90.42	90	0.469	99.58	97	0.408	85.99	90	0.601	87.99	93	0.628
REG27	68.27	76	0.725	89.02	80	0.229	66.2	74	0.730	70.41	77	0.690
REG3	68.21	57	0.147	74.36	58	0.073	65.29	55	0.161	68.86	56	0.116
REG3	66.65	75	0.744	67.54	77	0.771	57.5	75	0.934	57.73	76	0.941
REG3	54.66	78	0.979	62.31	80	0.929	61.97	77	0.894	63.64	80	0.910
REG3	88.66	84	0.343	90	82	0.255	92.28	82	0.205	96.72	79	0.086
REG4	99.28	100	0.502	102.71	105	0.546	97.23	101	0.588	100.43	100	0.470
REG4	62.05	74	0.838	69.08	76	0.701	62.94	74	0.817	62.27	74	0.833
REG4	107.3	108	0.501	107.02	109	0.536	111.48	109	0.416	117.04	108	0.259
REG4	114.36	107	0.295	116.83	113	0.383	111.38	109	0.418	114.13	106	0.277
REG4	97.65	80	0.087	97.88	83	0.126	97.14	79	0.081	101.28	78	0.039
REG5	96.94	99	0.541	96.61	100	0.578	96.3	101	0.614	96.22	100	0.589
REG5	172.4	118	0.001	190.72	117	0.000	166.73	116	0.001	211.43	112	0.000
REG5	128.92	99	0.023	128.17	99	0.026	126.35	98	0.028	130.93	98	0.015
REG5	95.89	89	0.290	93.91	89	0.340	95.55	88	0.273	101.26	86	0.125
REG5	129.62	83	0.001	128.37	83	0.001	133.95	84	0.000	144.96	82	0.000
REG6	79.36	84	0.623	79.18	83	0.599	81.35	81	0.469	75.8	79	0.582
REG6	92.38	88	0.353	92.98	92	0.453	86.59	88	0.523	90.3	88	0.412
REG6	116.75	94	0.056	115.56	93	0.057	113.04	91	0.059	119.85	90	0.019
REG6	96.25	91	0.333	97.66	92	0.323	96.58	91	0.324	95.18	91	0.361
REG6	90.18	72	0.072	95.18	75	0.058	93.66	72	0.044	93.73	73	0.052
REG7	86.36	100	0.833	94.73	96	0.518	88.07	98	0.754	93.36	96	0.558
REG7	125.78	79	0.001	127.33	80	0.001	124.16	79	0.001	124.78	79	0.001
REG7	98.29	76	0.044	105.8	75	0.011	89.17	75	0.126	93.2	75	0.076
REG7	105.4	86	0.076	108.89	85	0.041	97.36	84	0.151	97.25	83	0.136
REG8	78.46	67	0.159	76.44	67	0.201	71.65	67	0.326	72.07	67	0.314
REG8	70.06	55	0.083	67.3	55	0.123	63.58	52	0.130	62.56	52	0.150
REG8	57.18	72	0.899	61.95	73	0.819	57.07	70	0.867	62.03	74	0.839
REG8	75.42	81	0.655	75.73	80	0.615	65.29	80	0.883	62.08	78	0.907
REG9	87.81	124	0.994	88.18	123	0.992	78.51	122	0.999	85.39	118	0.990
REG9	117.13	115	0.427	126.28	110	0.137	104.02	109	0.617	117.95	106	0.201
REG9	102	107	0.619	101.39	107	0.635	91.35	103	0.788	92.19	103	0.769
REG9	128.52	108	0.087	123.53	107	0.131	113.01	105	0.279	127.21	102	0.046
REG9	111.49	90	0.062	110.5	90	0.070	104.32	88	0.113	115.56	87	0.022

	GRM			1PLGRM			GPCM			PCM		
Label	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability
FAR1	141.37	117	0.0621	149.52	112	0.0103	179.99	116	0.0001	196.1	111	0.0001
FAR1	166.45	133	0.0261	173.14	128	0.0049	171.57	134	0.0158	166.44	129	0.0147
FAR1	150.87	111	0.0071	158.34	111	0.0021	174.36	115	0.0003	164.65	110	0.0006
FAR10	125.73	106	0.0926	121.26	108	0.1805	116.89	106	0.2207	120.84	107	0.17
FAR10	109.46	99	0.2216	110.19	101	0.2496	110.28	99	0.2059	111.65	101	0.22
FAR10	106.33	108	0.528	116.64	113	0.3879	112.45	106	0.3152	124.09	114	0.2437
FAR10	90.54	115	0.9554	91.71	116	0.9533	91.69	111	0.9091	95.22	114	0.899
FAR10	116.02	108	0.2814	113.34	114	0.5006	117.75	109	0.2665	126.65	118	0.2763
FAR11	99.63	108	0.7055	99.7	110	0.7495	105.66	107	0.5191	103.65	107	0.5741
FAR11	121.73	123	0.516	124.65	121	0.3912	120.9	121	0.4861	128.21	118	0.2451
FAR11	134	116	0.1211	138.37	120	0.1203	129.7	115	0.1647	136.27	118	0.1197
FAR11	97.62	108	0.7535	98.44	109	0.7566	100.78	110	0.7244	102.14	107	0.6153
FAR11	136.96	136	0.4614	137.19	138	0.5039	125.46	135	0.7106	133.71	137	0.5641
FAR12	172.42	146	0.0667	178.4	150	0.0565	173.79	145	0.0516	172.07	147	0.0771
FAR12	174.56	143	0.0373	177.98	146	0.0368	169.34	140	0.0461	168.61	140	0.05
FAR12	168.58	149	0.1299	169.28	150	0.134	169.75	149	0.1172	167.54	149	0.142
FAR12	137.52	128	0.2666	150.35	139	0.2408	128.7	127	0.4417	128.29	129	0.5016
FAR12	143.32	139	0.3831	150.1	146	0.3907	145.31	139	0.3396	146.2	138	0.2999
FAR13	107.52	109	0.5229	107.61	111	0.5739	105.44	103	0.4144	109.47	107	0.4151
FAR13	106.78	109	0.5427	105.38	111	0.633	118.48	104	0.1569	117.48	107	0.2295
FAR13	105.95	107	0.5111	114.08	112	0.4271	101.48	102	0.4966	106.15	104	0.4223
FAR13	76.21	97	0.9414	96.55	103	0.6603	77.04	91	0.8516	88.46	100	0.789
FAR13	97.55	97	0.4659	112	100	0.1936	98.77	93	0.3211	108.19	97	0.2052
FAR14	141.23	118	0.0713	145.77	125	0.0986	144.08	116	0.0395	156.16	127	0.0403
FAR14	113.62	95	0.0934	136.98	103	0.0141	114.42	93	0.0652	131.34	102	0.0267
FAR14	145.07	107	0.0084	158.25	117	0.0067	151.38	110	0.0055	166.28	119	0.0028
FAR14	117.37	115	0.4205	141.29	126	0.1662	118.01	116	0.4301	145.42	126	0.1136
FAR14	122.99	101	0.0675	125.35	101	0.0506	124.77	99	0.0409	131.58	102	0.0259
FAR15	167.8	164	0.4029	168.38	164	0.3907	169.13	158	0.2578	173.69	153	0.1207
FAR15	147.77	161	0.7648	150.2	158	0.6589	156.01	160	0.5749	166.5	155	0.2495
FAR15	219.42	180	0.024	220.26	173	0.0088	203.59	178	0.0915	238.91	163	0.0001
FAR15	112.16	136	0.9331	116.28	136	0.8885	133.04	140	0.6494	176.84	127	0.0023
FAR15	140.36	173	0.9674	136.4	173	0.9817	158.54	173	0.7778	174.53	161	0.2202
FAR16	173.31	141	0.0333	167.23	136	0.0355	164.55	140	0.0765	175.82	127	0.0027
FAR16	148.65	151	0.5392	148.35	152	0.5689	150.8	149	0.444	158.14	145	0.215
FAR16	169.63	155	0.1993	160.47	152	0.303	171.3	153	0.1479	166.6	149	0.1537
FAR16	135.97	140	0.581	140.16	143	0.5521	143.36	143	0.4762	141.71	141	0.4679
FAR16	178.47	173	0.3715	178.79	173	0.3652	177.67	173	0.3876	185.5	167	0.1554
FAR17	197.66	149	0.0047	199.85	154	0.0076	193.15	144	0.0039	195.9	145	0.0031
FAR17	169.12	143	0.067	191.43	158	0.036	170.32	141	0.0467	171.98	146	0.0698

	GRM			1PLGRM			GPCM			PCM		
Label	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability
FAR17	133.79	134	0.4894	152.83	141	0.2339	123.46	132	0.6904	129.32	133	0.5744
FAR17	131.34	122	0.2654	148.67	134	0.1823	132.83	121	0.2176	135.29	125	0.2494
FAR17	143.54	125	0.1227	159.73	145	0.1903	151.97	129	0.0816	149.83	133	0.1509
FAR18	175.58	152	0.0924	183.76	145	0.0162	187.05	153	0.0317	211.31	143	0.0002
FAR18	116.66	140	0.9252	127.29	136	0.6914	131.11	142	0.7339	158.5	133	0.065
FAR18	147.49	143	0.3808	150.89	138	0.2137	159.2	143	0.1675	190.61	132	0.0006
FAR18	141.94	128	0.1884	136.86	126	0.2395	147.74	129	0.1238	162.5	117	0.0035
FAR18	126.67	123	0.3916	125.02	123	0.4319	123.91	119	0.3602	129.26	119	0.2449
FAR19	137.06	147	0.7105	161.67	163	0.5151	137.57	147	0.6997	137.54	147	0.7005
FAR19	124.21	129	0.6031	147.62	143	0.378	155.1	126	0.0401	148.29	126	0.0852
FAR19	141.24	148	0.641	163.48	159	0.387	121.98	140	0.8616	120.78	142	0.9012
FAR19	171.58	171	0.4737	190.65	185	0.3722	177.95	167	0.2664	183.15	170	0.232
FAR19	171.02	155	0.1792	189.83	168	0.1191	191.09	151	0.0151	191.17	151	0.015
FAR2	122.08	133	0.7417	123.46	135	0.7528	125.47	133	0.6665	133.12	129	0.3835
FAR2	142.81	155	0.7498	157.16	150	0.3277	130.05	157	0.9429	148.37	149	0.4996
FAR2	164.2	132	0.03	170.46	133	0.0157	155.4	130	0.0637	168.72	122	0.0033
FAR2	139.61	123	0.1451	141.26	130	0.2354	122.37	121	0.4487	120.68	124	0.5682
FAR2	140.22	125	0.1663	140.08	127	0.2015	123.08	125	0.5323	147.57	116	0.0254
FAR20	165.11	161	0.3954	165.23	163	0.4361	159.52	157	0.4286	160.29	160	0.4791
FAR20	129.37	132	0.549	134.71	136	0.5156	128.52	132	0.57	126.96	132	0.6081
FAR20	150.09	155	0.5967	162.72	166	0.5578	143.88	155	0.729	160.21	161	0.5033
FAR20	143.44	153	0.6988	147.61	163	0.8008	140.96	150	0.6897	140.19	158	0.8425
FAR20	139.53	146	0.6354	150.12	156	0.618	145.16	146	0.5045	151.86	150	0.4426
FAR21	134.6	124	0.2426	135.99	124	0.2174	137.5	125	0.2095	137.05	127	0.2556
FAR21	124.18	120	0.3779	131.98	125	0.3168	112.86	117	0.5915	118.75	121	0.5413
FAR21	139.55	119	0.0958	141.72	119	0.0761	145.14	118	0.0455	146.47	116	0.0293
FAR21	99.49	105	0.6339	94.67	107	0.7974	94.59	104	0.7348	93.3	103	0.7429
FAR21	104.22	125	0.9119	113.11	128	0.8233	100.88	125	0.9446	101.59	123	0.9209
FAR22	112.45	102	0.2251	117.17	103	0.1606	107.47	102	0.3358	110.02	105	0.349
FAR22	124.62	98	0.036	135.17	96	0.0052	131.25	96	0.0098	126.31	99	0.0334
FAR22	103.12	84	0.0767	99.5	87	0.1694	86.85	80	0.2808	111.72	86	0.0326
FAR22	121.63	101	0.0793	122.2	101	0.0742	111.46	102	0.2451	113.04	102	0.2136
FAR22	98.02	103	0.6206	105.16	98	0.2919	93.91	103	0.7281	92.51	104	0.783
FAR23	19.84	25	0.7558	48.9	23	0.0013	19.61	25	0.7678	19.58	25	0.7693
FAR23	38.63	26	0.0527	54.5	25	0.0006	38.85	26	0.0503	39.16	26	0.0469
FAR23	18.59	25	0.8166	28.4	24	0.243	18.57	25	0.8176	23.68	26	0.5955
FAR23	29.54	32	0.5926	105.25	27	0.0001	29.8	32	0.5792	31.85	32	0.4754
FAR23	26.97	31	0.6745	57.19	28	0.0009	27.36	31	0.6547	28.08	31	0.6183
FAR24	18.1	27	0.9007	55.69	25	0.0004	17.34	27	0.9225	17.4	27	0.9209
FAR24	25.34	25	0.4451	53.71	23	0.0003	25.38	25	0.4427	26.64	25	0.376

	GRM			1PLGRM			GPCM			PCM		
Label	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability
FAR24	51.41	27	0.0031	64.33	27	0.0001	40.85	27	0.0424	36.66	28	0.1262
FAR24	17.62	25	0.8584	46.59	24	0.0037	17.45	25	0.8653	17.4	25	0.8671
FAR24	25.94	27	0.5233	48.22	23	0.0016	25.76	27	0.533	27.13	27	0.4584
FAR25	37.33	28	0.1115	48.21	28	0.0101	37.57	28	0.1066	50.53	30	0.0109
FAR25	22.98	31	0.8501	38.91	29	0.1031	22.96	31	0.8508	33.74	33	0.433
FAR25	20.11	30	0.914	48.32	28	0.0099	20.2	30	0.9114	24.54	31	0.7886
FAR26	39.06	31	0.1511	49.8	31	0.0175	38.78	31	0.1586	38.82	31	0.1576
FAR26	36.84	27	0.0978	60.57	25	0.0001	35.3	27	0.1311	35.67	28	0.1509
FAR26	30.97	26	0.2288	62.39	23	0.0001	30.96	26	0.2291	31.8	26	0.1993
FAR26	26	27	0.5198	50.27	25	0.002	26.15	27	0.5116	28.28	28	0.4511
FAR27	24.08	29	0.7257	38.67	27	0.0678	24.35	29	0.7125	41.16	31	0.1046
FAR27	20.63	26	0.7614	23.23	26	0.621	20.66	26	0.7601	47.13	29	0.018
FAR27	36.06	24	0.054	37.17	25	0.0555	37.65	25	0.0499	77.84	27	0.0001
FAR28	18.46	25	0.8228	27.9	24	0.2637	18.41	25	0.8249	31.75	28	0.2841
FAR28	16.1	25	0.9119	20.72	26	0.7569	15.21	25	0.9365	31.27	28	0.3042
FAR28	36.06	29	0.1715	78.6	26	0.0001	34.23	29	0.2302	34.72	30	0.2525
FAR28	19.4	29	0.9107	54.86	26	0.0008	19.41	29	0.9105	19.42	29	0.9103
FAR29	115.99	103	0.1797	118.41	104	0.1579	102.14	101	0.4503	116.19	105	0.214
FAR29	110.87	109	0.4328	115.82	113	0.4083	111.36	109	0.4186	117.04	112	0.3528
FAR29	94.48	115	0.9191	94.88	114	0.9033	92.73	114	0.9283	94.37	112	0.8853
FAR29	114.81	99	0.1321	113.56	102	0.2039	112.05	99	0.1743	116.64	104	0.1868
FAR29	98.41	109	0.7572	100.25	111	0.7587	90.7	108	0.8852	93.29	110	0.8738
FAR3	139.1	139	0.4823	147.58	149	0.518	144.14	143	0.4582	146.89	144	0.4171
FAR3	169.82	128	0.0079	182.59	140	0.009	164.87	126	0.0114	166.3	132	0.0232
FAR3	138.01	130	0.2983	155.91	142	0.2005	140.52	129	0.23	145.2	139	0.3418
FAR3	114.35	122	0.6768	144.71	135	0.268	117.22	125	0.6778	122.92	127	0.5863
FAR30	148.61	147	0.4478	155.69	145	0.2571	166.92	147	0.1246	176.58	147	0.0485
FAR30	113.42	117	0.5769	114.04	117	0.5608	117.93	117	0.4593	120.79	116	0.3612
FAR30	127.24	106	0.0782	114.45	106	0.2703	137.67	106	0.0209	143.4	104	0.0063
FAR30	116.21	114	0.4243	116.11	117	0.5065	135.52	118	0.1288	133.23	123	0.2489
FAR31	102.54	99	0.3831	102.19	100	0.4197	92.7	93	0.4899	98.32	99	0.501
FAR31	78.41	96	0.9047	78.55	96	0.9026	81.19	96	0.8603	80.93	97	0.8803
FAR31	90.37	99	0.7209	90.46	99	0.7186	83.77	98	0.8468	85.45	100	0.85
FAR31	65.33	94	0.9893	67.15	95	0.9865	70.75	92	0.9512	70.14	92	0.9564
FAR4	97.95	113	0.8426	108.86	113	0.5931	107.18	118	0.7531	105.84	116	0.7404
FAR4	114.9	109	0.3305	114.32	109	0.3443	106.82	109	0.5417	108.24	109	0.5032
FAR4	108.53	121	0.7847	129.17	118	0.2268	104.39	120	0.8443	115.73	117	0.5164
FAR4	87.07	107	0.921	85.81	108	0.9432	95.28	107	0.7845	94.3	107	0.8051
FAR4	113.89	122	0.6877	116.86	123	0.6393	110.32	121	0.7471	111.76	121	0.715
FAR5	169.17	128	0.0086	168.19	127	0.0084	166.74	126	0.0088	160.77	121	0.0091

	GRM			1PLGRM			GPCM			PCM		
Label	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability
FAR5	146.34	141	0.3613	146.55	142	0.3791	141.93	139	0.4147	141.54	139	0.4236
FAR5	132.87	126	0.3199	135.02	130	0.3632	128	125	0.4086	127.61	125	0.4179
FAR5	163.56	165	0.5175	161.86	166	0.5766	163.76	160	0.4026	165.95	163	0.4207
FAR5	203.99	162	0.0141	190.18	156	0.0324	179.5	162	0.1644	187.09	160	0.0703
FAR6	144.03	163	0.855	143.94	158	0.7818	142.11	157	0.7971	159.8	152	0.3161
FAR6	147.11	145	0.4351	147.49	140	0.3154	151.64	142	0.2744	162.53	137	0.0673
FAR6	171.06	166	0.3774	163.01	153	0.2747	158.37	161	0.5442	171.96	147	0.0778
FAR6	144.36	143	0.453	160.34	143	0.1524	139.92	144	0.581	169.49	135	0.0237
FAR6	121.23	132	0.7394	121.06	132	0.7429	120.74	129	0.6861	120.09	130	0.7225
FAR7	108.37	108	0.4725	109.01	106	0.4006	101.34	108	0.6622	101.55	105	0.5778
FAR7	113.45	119	0.6267	143.63	114	0.0316	107.45	117	0.7254	136.43	112	0.0581
FAR7	160.73	118	0.0055	172.34	118	0.0008	159.65	117	0.0054	158.54	115	0.0045
FAR7	170.48	126	0.0051	187.98	120	0.0001	166.49	126	0.0091	181.66	119	0.0002
FAR7	174.72	145	0.0467	178.03	138	0.0123	163.99	141	0.09	158.17	139	0.1269
FAR8	134.19	142	0.6676	138.81	138	0.4653	123.97	139	0.815	127.61	129	0.5187
FAR8	172.95	143	0.0446	170.61	143	0.0574	177.52	139	0.0152	177.56	139	0.0151
FAR8	115.26	130	0.8187	114.76	132	0.8578	109.71	129	0.8898	102.97	125	0.9255
FAR8	162.95	172	0.6779	165.82	167	0.5116	154.95	169	0.7735	166.38	161	0.369
FAR8	139.78	150	0.7143	137.36	152	0.7969	148.27	151	0.548	166.19	140	0.0646
FAR9	125.45	122	0.3965	126.08	129	0.5569	116.78	120	0.5668	122.84	127	0.5883
FAR9	126.69	135	0.6832	128.86	139	0.7204	117.97	134	0.8366	142.27	143	0.5021
FAR9	123.77	119	0.3633	146.16	125	0.0948	120.27	117	0.3988	148.36	126	0.0846
FAR9	134.86	109	0.0471	163.31	123	0.0088	129.09	107	0.0718	154.43	121	0.0217
FAR9	152.32	126	0.0552	153.15	132	0.1004	145.76	124	0.0884	153.62	127	0.054

	GRM			1PL GRM			GPCM			PCM		
Label	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability	X^2	*d.f.*	Probability
AUD1	73.92	59	0.0912	73.24	58	0.0855	55.17	55	0.4691	59.89	57	0.3704
AUD1	47.03	43	0.3103	59.59	45	0.0712	47.72	42	0.2509	70.83	46	0.0108
AUD1	73.98	55	0.0447	66.72	55	0.1334	64.74	55	0.1728	68.56	57	0.1401
AUD1	75.64	56	0.0412	71.4	56	0.0804	56.11	53	0.3583	74.42	57	0.0604
AUD1	45.98	56	0.8283	57.29	58	0.5028	44.24	57	0.8916	64.53	61	0.3535
AUD10	121.07	115	0.3305	131.63	116	0.1521	119.08	113	0.329	121.99	111	0.2237
AUD10	159.25	167	0.6536	161.57	160	0.4508	158.03	166	0.6586	193.65	150	0.0094
AUD10	153.71	142	0.2368	153.18	142	0.2461	147.22	141	0.3425	168.72	134	0.0227
AUD11	119.7	107	0.1889	118.3	107	0.2139	115.66	106	0.2449	113.71	106	0.2864
AUD11	87.48	81	0.2913	94.96	87	0.2618	86.68	79	0.2592	94.46	85	0.226
AUD11	115.86	107	0.2624	117.54	109	0.271	112.63	104	0.2646	108.38	107	0.4451
AUD11	156.98	102	0.0004	172.13	107	0.0001	151.26	101	0.0009	165.31	105	0.0002
AUD12	105.45	106	0.4975	112.66	111	0.439	104.04	103	0.4536	105.02	108	0.5637
AUD12	127.19	118	0.2653	136.58	127	0.2646	137.76	119	0.1149	131.03	125	0.3377
AUD12	115.95	115	0.4584	112.5	115	0.5491	106.34	112	0.6336	109.36	111	0.5269
AUD13	106.2	78	0.0186	107.43	78	0.0152	107.54	77	0.0123	107.14	77	0.0132
AUD13	75.7	72	0.3594	86.44	69	0.0761	78.36	73	0.3123	80.45	69	0.1629
AUD13	90.44	71	0.0596	124.27	66	0.0001	89.84	70	0.0552	113.71	67	0.0003
AUD13	69	83	0.8651	95.46	80	0.1143	67.73	82	0.8717	80.38	81	0.4993
AUD14	111.72	131	0.8876	115.21	124	0.7019	109.46	127	0.8674	106.99	122	0.8318
AUD14	116.38	119	0.5513	118.58	118	0.4683	117.77	116	0.4374	118.69	115	0.3875
AUD14	100.09	91	0.2412	99.22	91	0.2603	94.69	88	0.2936	95.84	85	0.1976
AUD14	85.5	103	0.8943	89.65	102	0.8041	91.34	106	0.8443	89.65	97	0.6896
AUD15	97.27	100	0.5594	96.85	100	0.5713	100.25	97	0.3898	106.88	95	0.1902
AUD15	102.77	96	0.2993	100.38	100	0.4711	103.42	96	0.2839	109.99	95	0.1392
AUD15	121.56	109	0.1934	127.62	110	0.12	118.79	108	0.2246	120.54	108	0.1926
AUD16	116.06	90	0.0336	113.36	92	0.0647	115.14	90	0.0382	111.16	90	0.0646
AUD16	123.83	139	0.8174	129.99	138	0.6745	124.7	137	0.7663	147.08	130	0.1451
AUD16	119.14	123	0.582	117.88	119	0.5123	123.03	123	0.483	132.34	113	0.1031
AUD16	125.59	115	0.235	125.27	117	0.2835	124.72	114	0.2315	136.62	111	0.0498
AUD17	107.48	128	0.906	110.07	131	0.9079	94.09	125	0.9822	101.11	124	0.9347
AUD17	127.76	103	0.0495	133.51	104	0.0271	124.46	104	0.0836	115.76	101	0.1494
AUD17	139.08	135	0.387	139.02	135	0.3882	135.15	133	0.4312	127.08	127	0.482
AUD18	67.58	80	0.8378	72.29	77	0.6314	68.93	81	0.8285	76.01	80	0.6063
AUD18	92.77	77	0.1062	89.1	80	0.2274	95.14	76	0.0678	84.88	82	0.3913
AUD18	88.38	83	0.3222	89.09	83	0.3036	87.72	80	0.2594	86.01	81	0.3301
AUD19	102.91	82	0.059	102.01	82	0.0665	94.77	78	0.095	103.26	76	0.0205
AUD19	83.94	98	0.8435	86.89	102	0.8575	83.96	99	0.8602	83.97	99	0.86
AUD19	120.43	94	0.0344	122.18	95	0.0316	124.47	94	0.0194	123.26	91	0.0137

AUD19	77	91	0.8525	97.62	102	0.6048	84.58	92	0.6965	84.09	95	0.7813
AUD19	105.23	96	0.2435	111.09	99	0.191	107.62	98	0.2376	105.68	98	0.2799
AUD2	84.88	63	0.0345	91.39	58	0.0034	72.8	61	0.1429	72.84	58	0.0907
AUD2	74.29	64	0.1776	76.32	61	0.0892	65.73	62	0.3484	64.77	60	0.3134
AUD2	103.19	66	0.0023	102.61	66	0.0026	83.75	66	0.069	86.42	68	0.0651
AUD20	147.58	122	0.0573	148.18	122	0.0535	128.14	120	0.2885	136.03	114	0.078
AUD20	131.67	122	0.2589	144.87	128	0.1461	120.14	120	0.4799	123.04	120	0.4056
AUD20	122.87	114	0.2684	133.22	118	0.1599	130.78	113	0.1209	134.88	113	0.0785
AUD21	90.91	65	0.0186	86.74	63	0.0253	84.98	66	0.0578	76.91	63	0.1117
AUD21	60.04	60	0.4753	72.14	61	0.1553	59.68	61	0.5246	66.93	59	0.2232
AUD21	80.74	78	0.3928	95.55	76	0.0641	81.57	78	0.3684	80.3	77	0.3754
AUD21	83.85	72	0.16	83.65	71	0.1444	79.64	72	0.2508	73.36	70	0.368
AUD21	97.43	73	0.0296	97.99	72	0.0225	81.46	70	0.1644	82.33	69	0.1301
AUD22	84.97	86	0.512	85.57	91	0.6414	86.5	85	0.4352	89.21	88	0.4447
AUD22	67.93	59	0.1987	84.45	63	0.0369	72.69	57	0.0785	77.93	62	0.0832
AUD22	87.43	75	0.1541	98.14	79	0.0711	95.82	73	0.0378	96.46	77	0.066
AUD22	79.76	81	0.5188	87.28	87	0.4721	88.27	80	0.2463	85.95	84	0.4212
AUD22	104.18	87	0.101	111	93	0.0981	121.2	86	0.0074	116.33	91	0.0378
AUD23	81.3	87	0.6527	88.4	85	0.3784	86.52	88	0.5253	90.34	86	0.3528
AUD23	81.66	66	0.0924	85.42	65	0.0456	76.52	66	0.1762	83.29	64	0.0529
AUD23	85.74	83	0.3959	101.08	80	0.0557	80.79	81	0.4864	105.25	80	0.0307
AUD23	88.09	90	0.5379	98.67	88	0.2047	89.23	90	0.5037	92.13	86	0.3056
AUD24	40.09	23	0.015	23.42	22	0.38	22.55	22	0.4291	27.97	22	0.1762
AUD24	10.12	15	0.8127	13.89	15	0.5349	10.02	15	0.8188	22.95	16	0.1148
AUD24	13.15	19	0.8311	26.74	18	0.0838	13.4	19	0.8181	12.89	19	0.8448
AUD24	28.36	20	0.1008	26.74	20	0.1422	29.86	20	0.072	30.5	22	0.1065
AUD24	32.8	21	0.0484	47.36	19	0.0003	32.81	21	0.0482	32.31	22	0.072
AUD25	60.1	58	0.399	60.26	58	0.3932	68.4	58	0.1646	76.57	57	0.0428
AUD25	87.96	82	0.3057	86.85	85	0.4248	91.18	81	0.2057	87.77	83	0.3385
AUD25	85.74	84	0.4277	83.48	84	0.4964	84.48	80	0.344	83.62	82	0.4303
AUD25	88.03	75	0.1439	88.35	75	0.1388	89.33	76	0.1405	89.6	76	0.1362
AUD25	57.42	67	0.7922	60.96	67	0.6849	64.08	68	0.6131	64.42	65	0.4979
AUD26	78.6	70	0.2248	87.04	75	0.1611	83.34	70	0.1315	83.97	71	0.1391
AUD26	59.85	98	0.9992	60.23	101	0.9996	58.47	95	0.9988	57.29	100	0.9998
AUD26	110.61	97	0.1628	111.32	98	0.1686	108.14	97	0.2063	107.53	97	0.218
AUD26	78.5	73	0.3084	80.74	74	0.2763	67.59	71	0.5936	69.6	72	0.559
AUD26	91.81	95	0.5742	91.35	95	0.5875	91.68	94	0.5491	91.14	94	0.5648
AUD27	122.96	110	0.1875	131.93	111	0.0854	121.49	110	0.2133	121.37	110	0.2156
AUD27	144.52	126	0.1238	151.79	130	0.0928	132.82	123	0.2567	133.45	123	0.2447
AUD27	115.92	127	0.7502	117.42	126	0.6956	117.99	127	0.7047	126.25	118	0.2845
AUD27	138.3	110	0.0352	142.7	114	0.0355	121.4	110	0.215	125.33	110	0.1505
AUD28	114.74	108	0.31	127.93	113	0.1593	107.04	107	0.4813	108.93	109	0.4846
AUD28	96.28	110	0.8217	101.35	115	0.8145	94.78	108	0.8143	97.59	106	0.7083

AUD28	112.85	118	0.6172	118.84	125	0.6387	110.27	118	0.6815	111.13	119	0.6842
AUD28	136.85	123	0.1854	152.69	128	0.0674	136.35	122	0.1767	134.59	123	0.2236
AUD28	115.06	120	0.6107	120.6	125	0.5951	106.76	119	0.7822	110.42	120	0.7235
AUD29	20.9	19	0.3439	24.96	19	0.1615	20.74	19	0.3531	20.86	19	0.3462
AUD29	19.17	25	0.7897	36.12	24	0.0533	18.99	25	0.7981	23.02	26	0.6329
AUD29	30.09	22	0.1161	33	22	0.0617	30.47	22	0.1072	36.16	22	0.0291
AUD29	33.54	20	0.0293	44.67	19	0.0008	33.61	20	0.0288	34.11	20	0.0253
AUD29	25.17	21	0.239	35.76	20	0.0164	25.36	21	0.2313	27.44	21	0.1562
AUD3	80.58	69	0.1605	81.7	70	0.1598	85.46	68	0.0747	80.79	69	0.1566
AUD3	93.96	91	0.3944	94.12	92	0.4184	92.33	90	0.4117	95.94	91	0.3408
AUD3	93.55	85	0.2462	100.57	89	0.1887	86.25	83	0.3814	92.56	85	0.2692
AUD3	78.28	90	0.8066	84.55	91	0.6706	83.88	90	0.6621	79.93	93	0.8312
AUD30	29.35	22	0.1347	34.78	22	0.0408	28.78	22	0.1509	41.35	23	0.0108
AUD30	22.11	24	0.5741	35.56	22	0.0337	22.07	24	0.576	22.79	24	0.5337
AUD30	13.42	18	0.7665	15.58	18	0.6231	13.33	18	0.7723	35.15	18	0.009
AUD30	25.93	23	0.3033	39.38	22	0.0127	26.07	23	0.2968	29.43	25	0.2457
AUD30	24.01	22	0.3458	26.18	22	0.2432	23.48	23	0.4346	32.98	23	0.0812
AUD31	37.85	18	0.004	40.26	14	0.0002	43.14	18	0.0008	25.61	16	0.0596
AUD31	5.23	14	0.9825	7.13	13	0.8958	630.14	19	0.0001	5.38	14	0.9799
AUD31	21.65	15	0.1169	22.75	14	0.0643	21.58	15	0.1191	21.14	14	0.0978
AUD31	22.31	14	0.0722	20.42	13	0.0849	19.72	13	0.1022	24.21	14	0.043
AUD32	98.33	80	0.0802	95.01	82	0.1541	100.33	82	0.0825	96.44	83	0.1483
AUD32	77.85	71	0.2695	76.14	70	0.2872	73.99	69	0.3181	79.9	71	0.2194
AUD32	107.73	88	0.0751	105.68	87	0.0843	101.39	88	0.1555	101.96	86	0.1151
AUD32	66.75	84	0.9166	66.61	84	0.9186	70.54	81	0.7905	70.52	81	0.7911
AUD33	104.69	77	0.0196	103.1	75	0.0174	101.65	77	0.0314	102.77	78	0.0316
AUD33	74.15	81	0.6924	73.84	80	0.673	73.86	81	0.7009	73.41	79	0.6568
AUD33	153.38	109	0.0033	168.39	104	0.0001	146.93	108	0.0076	157.42	107	0.0011
AUD33	109.63	103	0.3086	112.32	99	0.1698	102.2	102	0.4764	103.13	102	0.4508
AUD33	120.35	95	0.0405	120.63	92	0.0242	121.67	96	0.0395	119.55	96	0.052
AUD34	23.04	24	0.5189	27.82	24	0.2672	23	24	0.5208	23.78	24	0.4754
AUD34	18.37	23	0.7378	19.45	21	0.5572	18.12	23	0.7518	26.01	25	0.4088
AUD34	15.51	24	0.9053	28.2	21	0.1342	15.77	24	0.8965	16.95	24	0.8514
AUD34	21.5	21	0.4307	26.89	20	0.1379	21.74	21	0.4162	24.4	21	0.2731
AUD34	35.45	21	0.0251	37.22	21	0.0158	35.33	21	0.0259	44.26	22	0.0033
AUD35	26.56	21	0.1855	31.28	21	0.069	28.19	22	0.1691	32.94	22	0.0625
AUD35	35.91	25	0.0728	47.35	22	0.0013	36.02	25	0.0711	38.26	26	0.0572
AUD35	13.01	22	0.9332	24.84	21	0.2536	12.88	22	0.9366	13.62	22	0.9146
AUD35	32.59	23	0.0882	46.63	22	0.0016	32.38	23	0.0923	32.52	23	0.0895
AUD35	36.66	22	0.0257	41.87	21	0.0044	35.63	22	0.0332	41.89	24	0.0132
AUD4	106.41	86	0.067	105.64	89	0.1099	98.88	85	0.1439	100.45	89	0.1909
AUD4	72.7	76	0.5866	72.12	78	0.6668	71.51	75	0.5935	74.65	76	0.5231
AUD4	76.99	66	0.1669	93.76	70	0.0305	80.11	66	0.1135	100.79	69	0.0075

AUD4	76.38	78	0.5315	80.99	84	0.5733	67.16	78	0.8049	94.17	85	0.2323
AUD4	81.49	77	0.3409	100.33	83	0.0946	72.36	76	0.5977	93.93	81	0.154
AUD5	42.49	77	0.9995	43.64	79	0.9996	47.62	79	0.998	46.92	78	0.998
AUD5	86.74	92	0.6357	103.84	90	0.1507	91.67	94	0.5494	92.09	90	0.4186
AUD5	102.22	88	0.1423	112.92	87	0.0323	101.01	88	0.1618	108.26	84	0.0386
AUD5	80.51	86	0.6472	85.14	84	0.4455	78.67	83	0.6146	81.69	83	0.5209
AUD5	98.14	88	0.2154	111.3	90	0.0634	101.22	89	0.1768	99.17	87	0.1752
AUD6	103.66	94	0.2324	111.44	99	0.1848	105.22	93	0.1817	104.89	92	0.1688
AUD6	114.57	120	0.6231	125.19	120	0.354	121.23	119	0.4255	130.07	114	0.144
AUD6	102.51	90	0.1729	104.51	90	0.1404	100.61	91	0.23	100.78	89	0.1848
AUD6	97.62	100	0.5492	111.11	109	0.4252	96.95	100	0.5684	100.79	105	0.5986
AUD6	124.64	120	0.367	124.65	120	0.3667	118.2	120	0.5299	112.23	115	0.5561
AUD7	53.36	60	0.7156	57.34	60	0.5744	55.13	62	0.7198	55.52	62	0.7071
AUD7	68.48	67	0.4278	99.53	65	0.0038	67.09	67	0.4749	78.13	64	0.11
AUD7	63.94	69	0.6504	66.01	66	0.4775	60.17	68	0.7396	59	65	0.6866
AUD7	81.78	70	0.1583	95.67	68	0.0151	77.07	69	0.2359	79.86	68	0.1536
AUD8	91.64	99	0.6881	101.4	105	0.5818	94.53	100	0.6361	92.89	100	0.6804
AUD8	144.57	98	0.0016	158.02	105	0.0006	146.57	98	0.0011	145.3	98	0.0014
AUD8	119.91	121	0.5114	128.25	126	0.4269	121.45	121	0.4719	121.37	120	0.4485
AUD8	138.78	113	0.0501	139.32	113	0.047	123.81	110	0.1736	124.3	107	0.121
AUD8	76.37	94	0.9079	92.04	97	0.6238	77.41	93	0.878	76.73	90	0.8397
AUD9	71.12	74	0.5741	70.28	72	0.5361	72.44	76	0.5951	68.01	73	0.644
AUD9	93.49	85	0.2473	94.46	86	0.2494	79.54	81	0.5257	75.64	78	0.5554
AUD9	85.33	89	0.5909	89.27	89	0.4728	82.6	89	0.6709	80.39	89	0.7317