



Published in final edited form as:

Appl Psychol Meas. 2009 September 1; 33(6): 419–440. doi:10.1177/0146621608327801.

Comparison of CAT Item Selection Criteria for Polytomous Items

Seung W. Choi and

Northwestern University Feinberg School of Medicine

Richard J. Swartz

University of Texas M. D. Anderson Cancer Center

Abstract

Item selection is a core component in computerized adaptive testing (CAT). Several studies have evaluated new and classical selection methods; however, the few that have applied such methods to the use of polytomous items have reported conflicting results. To clarify these discrepancies and further investigate selection method properties, six different selection methods are compared systematically. The results showed no clear benefit from more sophisticated selection criteria and showed one method previously believed to be superior—the maximum expected posterior weighted information (MEPWI)—to be mathematically equivalent to a simpler method, the maximum posterior weighted information (MPWI).

Keywords

computerized adaptive testing; item response theory; polytomous models

Item selection, a core component of computerized adaptive testing (CAT), can enhance test measurement quality and efficiency by administering optimal test questions or items. The topic of item selection has received substantial attention in the measurement literature, and a number of selection techniques have been proposed (see van der Linden, 1998, and van der Linden & Pashley, 2000, for a combined comprehensive review). In its pure form (unconstrained item-level adaptation), CAT selects items sequentially to minimize the standard error of the current theta estimate. Classical selection methods either maximize Fisher's information (MFI) at the interim theta estimator (Lord, 1980; Weiss, 1982) or minimize the posterior variance of the estimator (Owen, 1975; Thissen & Mislevy, 2000; van der Linden & Pashley, 2000).

Alternative selection methods using a more global consideration of information have been proposed and applied to dichotomous items. Veerkamp and Berger (1997) demonstrated that selecting items based on integrating the information function weighted by the likelihood of theta, known as the maximum likelihood weighted information (MLWI), is superior to the traditional maximum point information criterion. They also showed that selecting items using the traditional MFI criterion evaluated at the expected a posteriori (EAP) estimate (instead of the maximum likelihood estimate) could realize a gain in efficiency similar to that found using the MLWI. With smaller mean squared errors (Bock & Mislevy, 1982; Wainer & Thissen, 1987), EAP as an interim theta estimator can also reduce the likelihood of optimizing at the wrong places on the theta continuum at the early stages of CAT. Chang and Ying (1996) and Chen, Ankenmann, and Chang (2000) explored the Kullback-Leibler (KL) information and

compared it to both MFI and MLWI. Studies found that global measures performed similarly and slightly better than the MFI, when the test length was less than ten items (Chen et al., 2000; Fan & Hsu, 1996).

van der Linden (1998) and van der Linden and Pashley (2000) studied some criteria that use Bayesian concepts to create global information. The minimum expected posterior variance (MEPV; also discussed in Owen, 1975), the maximum expected information (MEI), and the maximum expected posterior weighted information (MEPWI) use the posterior predictive probability distribution to average over predicted responses to the next item. The MEPWI also incorporates the posterior distribution (details forthcoming). A simpler criterion, the maximum posterior weighted information criterion (MPWI) used the posterior distribution to weight the information function. Using dichotomous items, van der Linden (1998) and van der Linden and Pashley (2000) reported that MFI performed the worst and the MEI, MEPV, and MEPWI performed similarly and significantly better than the MPWI. Also, for moderate length tests (ten items), they found little difference between the MPWI and the MFI criterion. They concluded that weighting the information function with just the posterior distribution of θ had little benefit.

In one of the first studies exploring polytomous items, van Rijn, Eggen, Hemker, and Sanders (2002) compared the maximum interval information (a.k.a. the A-optimality criterion [Passos, Berger, & Tan, 2007]) and the traditional MFI criteria under the generalized partial credit model (Muraki, 1993) but reported little difference in performance. However, the tests administered a large number of items due to extreme stopping rules. Thus, potential advantages may have been masked (Chen et al., 2000; van der Linden & Pashley, 2000).

Penfield (2006) examined the MEI and the MPWI under a polytomous model (the partial credit model). He showed that the MEI and MPWI performed similarly and were associated with more efficient trait estimation compared to the MFI. This contrasts with previous results finding that the MPWI performed worse than MEI. Penfield also conjectured that the MEPWI would be superior to the MEI or MPWI because the MEPWI attempts to combine the best of the two selection criteria.

Such conflicting findings of performance between the newer item selection methods versus the classical MFI inspired us to undertake this study. Furthermore, interest in polytomous items is growing with the recent use of CAT technology in patient-reported outcomes (PROs) such as mental health, pain, fatigue, and physical functioning (Reeve, 2006). Most PRO measures are constructed using Likert-type items more befitting of polytomous models. The advantage of using CAT in PRO measurement is that it reduces patient burden while achieving the same or better measurement quality, and it offers real-time scoring and electronic health status reporting. Our purpose is to clarify the discrepancies and further investigate the properties of various selection criteria under polytomous CAT in a health-related quality of life (HRQOL) measurement setting.

Method

For this study, we compared CAT under the graded response model (Samejima, 1969) using the following item selection procedures: maximum Fisher information (MFI), maximum likelihood weighted information (MLWI), maximum posterior weighted information (MPWI), maximum expected information (MEI), minimum expected posterior variance (MEPV), and maximum expected posterior weighted information (MEPWI). Because exposure and content control is less of an issue in PRO measurement than in high-stakes testing (Bjorner, Chang, Thissen, & Reeve, 2007), our focus is on the measurement-centered item selection component of a CAT.

Fisher's and Observed Information Measures

The distinction between Fisher's information and the observed information—that Fisher's information is the expectation (taken with respect to the responses) of the observed information—was noted by van der Linden (1998) and van der Linden and Pashley (2000). For Bayesian selection criteria that take the expectation over predicted responses, they proposed that the observed information measure is a more appropriate choice. The two information measures, however, become equivalent under certain item response theory (IRT) models that belong to an exponential family (Andersen, 1980). Specifically, van der Linden noted the mathematical equality of the two measures under dichotomous logistic models (one- and two-parameter logistic models).

It is also straightforward to show that the two information measures are identical under the generalized partial credit model (Donohue, 1994; Muraki, 1993). However, this equality does not hold for the graded response model (Samejima, 1969) with three or more response categories. This is evident when comparing the negative second derivative of the log-likelihood with the expectation of the same function over the responses (Baker & Kim, 2004). Thus, when using the graded response model, Bayesian item selection criteria based on Fisher's information and the observed information are not guaranteed to perform similarly (especially for the first few items). We chose to explore the difference between these two information measures under the graded response model in the selection methods that use observed information, namely the MEI and MEPWI.

Bayesian Selection Criteria

Adopting the notation from van der Linden and Pashley (2000), let R_k represent the set of items that are candidates for selection at stage k and let u_{ik} represent the response to the item i_k given at the k th stage in the sequence. Let $g(\theta|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}})$ denote the posterior distribution, which is the probability distribution of the latent trait of interest, θ , given the previous $k-1$ item responses. Let $p_j(r|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}})$ denote the posterior predictive distribution, which is the probability of giving response r to item j given the previous response history. For more details about these distributions, see the work of van der Linden and Pashley (2000), Carlin and Louis (2000), or Gelman, Carlin, Stern, and Rubin (2004). Also let $J_{u_{i_1}, u_{i_2}, \dots, u_{i_k}}(\theta)$ denote the observed information function and $I_{U_{i_1}, U_{i_2}, \dots, U_{i_k}}(\theta)$ denote Fisher's information function for items i_1, i_2, \dots, i_k . As mentioned in the introduction, the relationship between the two is

$$I_{U_{i_1}, \dots, U_{i_2}, \dots, U_{i_k}}(\theta) \equiv E \left[J_{U_{i_1}, \dots, U_{i_2}, \dots, U_{i_k}}(\theta) \right], \quad (1)$$

where the expectation is taken over the responses.

The MPWI for polytomous models is defined as

$$i_k \equiv \arg \max_j \left\{ \int g(\theta|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) I_{U_j}(\theta) d\theta : j \in R_k \right\}, \quad (2)$$

The selected item, i_k is the item remaining in the item bank that maximizes the integral of Fisher's information function weighted by the posterior distribution. As with the MLWI, this is a more global measure of information because the integral accounts for a range of theta values and the posterior distribution weights the values according to their plausibility given the previous item responses.

The maximum expected information (MEI) criterion extended to polytomous models is

$$i_k \equiv \arg \max_j \left\{ \sum_{r=1}^{m_j} p_j(r|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) J_{u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}, U_{j=r}} \left(\hat{\theta}_{u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}, U_{j=r}} \right) : j \in R_k \right\}, \quad (3)$$

where the summation is over the number of possible response categories for item j , m_j . The item i_k to be selected as the k th item is the one that maximizes the observed information we expect from each item j given the previous responses. That is, for each possible response r in $\{1, 2, \dots, m_j\}$ to a given item j , the observed information function associated with response r to item j is evaluated at the theta estimate based on previous responses and response r : $\hat{\theta}_{u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}, U_{j=r}}$. Since the actual response has not been observed, the expectation is taken over all responses to item j with respect to the posterior predictive probability distribution. Note that because the MEI depends on the observed information, we also generate results for the MEI whereby the observed information is replaced with Fisher's information for the next item to explore the potential effects.

Replacing the observed information function with the posterior variance yields the minimum expected posterior variance (MEPV) or maximum expected posterior precision criterion as follows:

$$i_k \equiv \arg \max_j \left\{ \sum_{r=1}^{m_j} p_j(r|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) \text{Var}(\theta|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}, U_{j=r}) : j \in R_k \right\}, \quad (4)$$

where $\text{Var}(\theta|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}, U_{j=r})$ is the posterior variance for item j with predicted response category r . This objective function is the Bayesian counterpart of the maximum information (Thissen & Mislevy, 2000) and hence provides a theoretical justification under the EAP. Also as a characteristic of the posterior distribution for theta, the variance can be considered a more global measure of information than Fisher's information.

Also proposed by van der Linden and Pashley (2000) was a variation of the MEI, called the maximum expected posterior weighted information (MEPWI) criterion. This criterion, in addition to taking the expectation over the possible responses to the proposed item, averages the expectation of the observed information function for each response over the entire theta range using the posterior distribution that includes the potential new response, r .

$$i_k \equiv \arg \max_j \left\{ \sum_{r=1}^{m_j} p_j(r|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) \int g(\theta|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}, U_{j=r}) \times J_{u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}, U_{j=r}}(\theta) d\theta : j \in R_k \right\}. \quad (5)$$

Equivalence of MEPWI and MPWI

On closer inspection, the MEPWI produces identical results regardless of whether the observed information or Fisher's information is used (even under the graded response model). More important, the MEPWI and the MPWI are mathematically equivalent for any parametric IRT model with a defined second derivative of the log-likelihood, including dichotomous IRT models. Because the items are assumed locally independent given theta, taking the expectation of the observed information with respect to both the predictive probability distribution and the posterior distribution based on the predicted response is the same as taking the expectation of Fisher's information with respect to the posterior distribution. A proof of both characteristics

is included in the appendix. Because of this equivalence, simulation results using the MEPWI are not presented.

Comparison Studies

We used a software program (Firestar; Choi, in press) to conduct CAT simulations under three fixed-length conditions (5, 10, and 20 items) using the EAP theta estimator and a standard normal prior. We selected the three test lengths to represent short, medium, and long tests under polytomous CAT. For benchmarking purposes, we used two naive item selection criteria as the potential lower and upper bounds, respectively: (a) a method selecting the next item at random, and (b) a method selecting the next item according to the MFI but using the known theta value (with simulations) or the EAP theta estimate based on the entire bank (with real data) instead of an interim theta estimate. We then constructed static forms of the same lengths (i.e., 5, 10, and 20 items) by selecting items with the largest expected information where the expectation was taken over the standard normal distribution (Fixed). We designed four main studies: Study 1 involved simulations based on real data and real items; Study 2 involved simulated data from the real items in Study 1; and Study 3 and Study 4 involved simulated data from two different simulated item banks (described below).

Study 1: Real Bank and Real Data

We obtained response data on a bank of 62 depression items from a sample of 730 respondents who had been patients or caregivers at M. D. Anderson Cancer Center in Houston, Texas. We then fitted the items with the graded response model (Samejima, 1969). Each item had four response choices, “Rarely or none of the time,” “Some or a little of the time,” “Occasionally or a moderate amount of time,” and “Most or all of the time.” All items retained in the bank had satisfactory fit statistics ($S-X^2$; Orlando & Thissen, 2003). Figure 1 shows the scale information function and the trait distribution. The scale information function is shifted markedly to the right in relation to the trait distribution. This occurred for two reasons. First, it is very difficult to generate items targeting the lower range for the construct of depression. Consequently, there are plenty of items measuring moderate to severe levels of depression but very few items measuring the lower end. Second, the sample was drawn from a general population as opposed to a clinically depressed population. The vertical hash marks above the horizontal axis of the graph denote the locations of the category boundary parameter estimates. The lowest value across all items was around -1.0 , implying that the measurement of examinees below this level would be inevitably poor (i.e., floor effects). As a screening tool, however, the primary interest of a measure like this lies in the upper region (e.g., > 1.0), and hence the lack of coverage of the healthier end of the continuum may not pose a problem. We conducted post-hoc CAT simulations using the calibration data ($N = 730$).

Studies 2, 3, and 4: Simulated Data and/or Simulated Item Banks

For Study 2, we used the same item bank as in Study 1; however, Study 2 used known theta values and responses simulated from them, as described below. For Studies 3 and 4, because our real item bank did not cover the full range of the trait spectrum, we simulated two different item banks of 65 items. We generated one bank to have “peaked” items and a second bank to have “flat” items, while maintaining similar levels of total information. Modeling after the item parameter estimates from the real bank, the peak bank was generated to contain items with relatively high discrimination parameters and category thresholds positioned close to each other. The flat bank was generated so that items on average had lower discrimination parameters and category thresholds that were spread wider apart.

Simulee Generation

To generate simulee responses for the real and simulated item banks, we generated a sequence of theta values between -3 and 3 with an increment of 0.5 . We then used the graded response model to simulate 500 replications of response patterns at each theta value for a total of 6,500 simulees over the entire sequence.

CAT Implementation

The standard normal distribution was chosen as the prior for θ for the expected a posteriori (EAP) estimation and all Bayesian item selection criteria. For the MFI, the first item was selected to maximize Fisher's information at the mean of the prior distribution (i.e., 0.0). We used a uniform likelihood function to select the first item for the MLWI and the standard normal prior distribution to select the first item for the MPWI.

Results

Study 1: Real Bank and Real Data

Table 1 shows the mean squared standard error of measurement (SE^2), root mean squared difference (RMSD) and correlation (CORR) between the theta estimates based on CAT and the full bank (i.e., using all 62 items) for all the selection methods and all test lengths. Except for the poor performance of the random selection (Random) and the static form (Fixed), no criterion clearly distinguished itself as better than the others. Overall, the differences between all the methods were small, with smaller differences exhibited in the longer test conditions. Figure 2 summarizes the standard error of measurement (calculated as posterior standard deviation) for the theta estimates from the five-item, real bank, real data conditions. The scatterplots show the posterior standard deviations against the CAT theta estimates. The dotted lines correspond to the conventional reliability of roughly 0.9, 0.8, and 0.7, respectively, assuming theta is distributed $N(0,1)$ in the population. At the bottom of each scatterplot is the proportion of the standard error estimates that are at or below 0.32 (corresponding to a reliability of 0.9). The proportions were very similar to each other, ranging from .53 to .55 for the six selection criteria, namely, MFI, MLWI, MPWI, MEI(Fisher), MEI(Observed), and MEPV. It is worth noting that the six selection criteria performed as well as the CAT that selected items maximizing information at the full-bank theta estimates (as seen by comparing the other plots to the plot labeled *Theta* in Figure 2). Also, the static short form performed substantially worse than CAT (see Table 1 and Figure 2). The static short form appeared to have selected items primarily from the center of the trait distribution (and the scale information function) and performed the worst. A similar but less pronounced pattern was observed when ten items were used (ranging from .73 to .76). This trend continued when 20 items were used (ranging from .86 to .87; further data not shown).

A somewhat noticeable degradation in performance was observed with the static fixed form when 5 and 10 items were used (.49 and .64). As evident in Figure 2, the theta estimates from the static fixed form were markedly degenerative (i.e., had fewer possible estimated values with higher SEM) as the estimates deviated from the center of the test information function (i.e., around $\theta = 1.0$, see Figure 1).

We also examined the percentage of overlap of selected items between the MFI selection method and each of the other selection methods, namely, MLWI, MPWI, MEPV, MEI(Fisher), and MEI(Observed). The overlap proportions were high (ranging from .84 to .90), even when using five items. The overlap increased when 10 items were used (.90 to .94 overall), and the overlap increased further when 20 items were used (.96 to .98).

The results from Study 1 reflect a specific setting and are largely dependent on the item bank characteristics. The item information functions in the real bank overlap substantially with each other (data not shown), leaving little room for the advanced selection criteria to demonstrate potential advantages. However, using simulated banks (Study 3 and Study 4) also provided similar results.

Studies 2, 3, and 4: Simulated Data, Real and Simulated Banks

Table 2 and Table 3 show the bias, root mean squared error (RMSE), and correlations (CORR) with the true theta values from CAT simulations using simulated responses generated from the real and simulated item banks, respectively. The results were similar to those from Study 1 in that no method distinguished itself as better than others, and overall the differences between the methods were small, with smaller differences exhibited when longer test conditions were used. Also, except for the benchmark random item-selection method, all three test conditions (5, 10, and 20 items) had high correlations (above .95) with the true theta values. The random selection method produced a correlation less than .95 only when five items were used ($r = .94, .92, \text{ and } .93$, for real, peaked, and flat banks, respectively).

Although the outcome measures for all the different CAT item selection methods were very close, all methods performed better on the bank of peaked items compared to the bank of flat items in terms of correlation and the standard error of measurement. However, smaller absolute biases were associated with the flat item bank. All CAT item selection criteria outperformed the static short form and the benchmark random item-selection method under all of the test lengths and bank conditions. Also, except for when five items were used, all CAT selection criteria perform very similarly. When five items were used, there was some slight separation between the methods, but there was no clear “winner.”

We again examined the item overlap between each of the methods—MLWI, MPWI, MEPV, MEI(Fisher), and MEI(Observed)—and the MFI selection method. In the five-item condition for the peaked and flat banks, the overlap between the MLWI criterion and the MFI was less than the other criteria. In the peaked bank, the overlap proportion between MLWI and MFI was .42, whereas the proportions were greater than .90 for other selection criteria. In the flat bank, the disparity between the overlap rates reduced; the MLWI and the MFI overlap rate was .73 versus .94–.99. In the real item bank, the MLWI overlap rate was similar to the others, all falling within .77 and .87. Similar patterns emerged in the 10- and 20-item conditions with substantially increased overlap proportions.

Conditional Results

Figure 3 through Figure 6 show various performance measures across the theta range for the different selection methods in the five-item, simulated data condition. Figure 3 shows the simulation results for the real item bank; Figure 4 shows results for the peaked item bank, and Figure 5 shows results for the flat item bank. As was the case in the study using the real bank and real data (Study 1), we found that the results from the various methods essentially converge as the test length increases. Thus, data from simulations using 10 and 20 items are not shown. Even when five items were used, in all the plots shown, most of the CAT selection methods performed very similarly to one another across the theta range.

Figure 3 through Figure 5 show the mean squared error (MSE) across the theta range for all the CAT selection methods, the static short form, and the two benchmarking criteria for the real, peaked, and flat banks, respectively. Figure 4 shows some striking behavior for the MLWI and the short form in the peaked bank. The MLWI had a much higher RMSE (.48) and had a slightly higher bias in the lower extremes of the theta distribution compared to the other methods (discussion forthcoming). The static short form performed well at the middle and

worse at the extremes. This occurs because the standard normal distribution heavily weights the middle range of theta values.

The conditional bias behaved as expected—values below the mean were overestimated, and values above the mean were underestimated. The bias increased in magnitude as one moved away from the population mean (zero) toward the extremes, especially when the item banks had low coverage at the extremes. The bias had the highest magnitude at the lower extreme for the real item bank. Also the static short form and the benchmarking random item selection method exhibited more bias—especially for theta values away from the mean—than any of the other CAT methods.

Discussion

A few limitations of this study are worth mentioning. First, we focused on the unconstrained CAT where the selection of items was solely based on the psychometric properties of items, and neither exposure control nor content balancing was considered. In health-related quality of life (HRQOL) measurement settings, content and exposure control are not necessarily of interest and thus were not explored in the current study. Second, the Kulback-Leibler information methods were not explicitly considered. Based on findings in the literature, we speculated that they would perform similarly to the other criterion that take a more global approach to the information, but further study may be helpful. Third, in accordance with most CAT studies in the literature, we used a standard normal prior distribution for theta when applying Bayesian methods. Although theta is often assumed to follow a standard normal distribution, this assumption may be untenable for some psychopathology or personality constructs. Future research might investigate robustness or sensitivity of these methods to misspecified priors.

Under the conditions considered, our study showed that the MFI (with the EAP interim theta estimator), the MLWI, and the Bayesian CAT selection methods in general performed equally well across all conditions. In addition, the MPWI and the MEPWI methods were determined to be mathematically equivalent. As expected, the worst performances we observed were those using the short form selection criterion. These results, however, may not be generalized to all short-form selection criteria and may apply only to short forms generated by selecting items that maximized the expected information over a standard normal distribution.

It is interesting to note that item selection based on the interim theta estimates performed equally as well as item selection based on the true theta. As a matter of fact, item selection based on the true theta did not appear to perform the best when compared to the other criteria. Specifically, the RMSE was larger when compared to other methods, albeit the difference was minute. A couple of explanations are tenable. First, the item banks had low coverage at the extreme theta levels where items selected based on the true theta were in general very poor items—flattened information functions with somewhat obtuse tail ends. This would inflate the conditional variance of theta estimates and increase the overall RMSEs. Second, other methods based on estimated theta values picked at least one item (i.e., the first) providing maximal information around the prior mean, which would produce more consistent (albeit biased) estimates due to the lack of informative subsequent items for the extreme theta values.

In the peaked bank, the MLWI had noticeably poorer performance at the lower theta values. To further explore this, we plotted the SEM versus the estimated theta values from simulated data using the peaked item bank (Figure 6). When using five items, we found that the MLWI method clearly had trouble estimating theta values between -2.0 and 0.0 . The SEM seemed to be higher in that region, and there were few different estimates for simulees with true theta values in that region. On further inspection of the peaked item bank, we noticed that the five

items with the highest information were in the higher theta range, and the item with the most information overall was located at 3.3. This finding explained the behavior of the MLWI. Because the likelihood function is conventionally set to be uniform before the first item is given, the MLWI method selected the item that had the largest area under the information function (with location 3.3) as the first item. Using a short test length, the estimates for the simulees at the negative end of the distribution had a higher degree of bias even after four additional items were administered as a result of the extremely inappropriate first item. Although such an item may be rare in practice, it does show a potential limitation of the MLWI. Some sort of heuristic must be set to select the first item, just as with the MFI, because technically the likelihood does not exist until the first item is selected.

In addition, this study noted that the graded response model (GRM) is the only model considered thus far in the literature for which the expected and observed information functions are not the same. However, from the results in the present study, we learned that the MEI using the observed information function behaved similarly to the MEI using Fisher's item information function. This suggests that when using simulated data sets, there is no practical difference between using the MEI based on Fisher's item information and using the MEI based on the observed information under the GRM.

We made an important observation in this study. As demonstrated mathematically in the appendix, the MEPWI, which was believed to be superior to MFI and MPWI, is actually equivalent to the much simpler and more straightforward MPWI criterion. This is an encouraging finding because the conceptual advantage of combining two beneficial approaches is actually realized by simply weighting the information function by the posterior distribution.

The results in our study, combined with the previous results of Chen and colleagues (2000) and Penfield (2006), suggest that selecting an item based on summarizing the information for a range of theta values (i.e., using a weight function such as the posterior distribution) as opposed to the information at a single point on the theta range might be all that is necessary to achieve the performance gain, and predicting the next observation using the predictive posterior distribution may not be necessary. In fact, this is also congruent with the findings of Veerkamp and Berger (1997) when they noted that MFI with the EAP estimate produces similar gain to the weighted information functions. This also may partially explain why all methods explored in our study performed similarly; if an estimate of theta was necessary, the EAP estimate was always used, even with the MFI selection method.

In summary, for polytomous items, the advantages that many of the Bayesian selection methods exhibited in the dichotomous setting might be masked in many practical settings using polytomous models unless the item bank is huge and the item information functions cover narrow ranges. For psychological and health outcome measures, this is seldom the case. Typical item banks for these constructs tend to have a small number of four- or five-category polytomous items (e.g., 30 to 60). Thus, their information functions can be multimodal (Muraki, 1993) and tend to span a wide range. Item selection is then mostly governed by the slope parameter instead of the category parameters. Furthermore, our simulations showed that for polytomous items modeled with the GRM, more complex and computing-intensive item selection procedures (e.g., MEI) with theoretical advantages do not seem to provide practical benefits over the standard maximum information criterion in conjunction with the EAP. The MFI using the EAP theta estimation was very competitive (although the MEPV method would be preferred from the Bayesian perspective). These findings suggest that, for item banks with a small number of polytomous items, any of the methods considered in this study are appropriate.

Acknowledgments

This research was supported in part by an NIH grant PROMIS Network (U-01 AR 052177-04, PI: David Cella) for the first author; and in part by an NCI career development award (K07-CA-113641, PI: Richard Swartz) for the second author. Collection of the data used in this study was funded in part through a core facility of a Cancer Center Support Grant (NCI CA16672, PI: John Mendelsohn). The authors would like to thank Gregory Pratt for his contribution to the literature search; Dennis Cox, Gary Rosner, and Wim van der Linden for their contributions through informal discussions; and Lee Ann Chastain, Lorraine Reitzel, Michael Swartz, two anonymous reviewers and the editor for their comments and critiques of the manuscript.

Appendix

Let $P_{ir}(\theta)$ denote the probability for responding in category $r \in \{1, 2, \dots, m_i\}$, where m_i is the maximum number of categories for item i . Also let item i_k represent the item administered at stage k . So u_{i_k} denotes the response to the item administered at stage k . Let $g(\theta|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}})$ be the posterior at stage $k-1$, and let $p(U_{i_k} = r|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}})$ be the posterior predictive distribution predicting the response to item i_k to be r given the previous $k-1$ observations as defined by van der Linden (1998) and van der Linden and Pashley (2000).

The general form of the likelihood, $L(\theta|\mathbf{u})$, for n items with responses $\mathbf{u} = (u_1, u_2, \dots, u_n)$ with $u_i \in \{1, 2, \dots, m_i\}$ categories is

$$L(\theta|\mathbf{u}) = \prod_{i=1}^n P_{u_i}(\theta) = f(\mathbf{u}|\theta). \tag{A1}$$

The *observed information function* for items i_1, i_2, \dots, i_k , $J_{u_{i_1}, u_{i_2}, \dots, u_{i_k}}(\theta)$, is defined by van der Linden (1998) as the negative second derivative of the log of the likelihood, and *Fisher's information*, $I_{U_{i_1}, U_{i_2}, \dots, U_{i_k}}(\theta)$, is defined as the expectation of the observed information function with respect to the data as given in Equation 1.

To simplify the notation, $I_{U_j}(\theta)$ denotes Fisher's information for a single item j , and $J_{U_j=r}(\theta)$ is the observed information evaluated for one item j with response $U_j = r$.

The MPWI is defined by van der Linden (1998) as shown in Equation 2 and the MEPWI as shown in Equation 5. We denote the MEPWI as shown in Equation 5 as *MEPWI(Observed)* because it uses the observed information function. *MEPWI(Fisher)* represents the MEPWI with Fisher's information replacing the observed information. To begin, notice that in Equation 5, the observed information function can be considered the sum of individual observed information functions: $J_{u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}, U_j=r}(\theta) = J_{u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}}(\theta) + J_{U_j=r}(\theta)$. Rearranging terms and separating the integral into the sum of two integrals yields

$$\begin{aligned} \text{MEPWI}(\text{observed}) \equiv & \arg \max_j \left\{ \underbrace{\int J_{u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}}(\theta) \sum_{r=1}^{m_j} p(U_j=r|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) g(\theta|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}, U_j=r) d\theta}_C \right. \\ & \left. + \int \sum_{r=1}^{m_j} p(U_j=r|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) g(\theta|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}, U_j=r) J_{U_j=r}(\theta) d\theta : j \in R_k \right\}. \end{aligned} \tag{A2}$$

Consider the first term in equation A2 (labeled C). By the definition of conditional probabilities, the product of the predictive probability distribution and the posterior distribution is the joint distribution of θ and the next observation, $g(\theta, U_j = r|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}})$. Taking the sum of this

joint distribution over the responses, r , to the next item results in the posterior distribution conditional on only the previous $k-1$ observations:

$$\begin{aligned}
 C &= \int J_{u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}}(\theta) \sum_{r=1}^{m_j} p(U_j=r|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) g(\theta|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}, U_j=r) d\theta \\
 &= \int J_{u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}}(\theta) \sum_{r=1}^{m_j} g(\theta, U_j=r|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) d\theta \\
 &= \int J_{u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}}(\theta) g(\theta|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) d\theta,
 \end{aligned} \tag{A3}$$

so that C is constant with respect to the maximization (i.e., it no longer depends on the next item) and can be dropped.

Next consider the second term from Equation A2. Notice that, again, the product of the predictive probability distribution and the posterior distribution is just the joint distribution of θ and the next observation, so continuing from Equation A2 we have

$$\begin{aligned}
 &= \arg \max_j \left\{ C + \int \sum_{r=1}^{m_j} p(U_j=r|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) g(\theta|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}, U_j=r) J_{U_j=r}(\theta) d\theta : j \in R_k \right\} \\
 &= \arg \max_j \left\{ \int \sum_{r=1}^{m_j} p(U_j=r|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) g(\theta|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}, U_j=r) J_{U_j=r}(\theta) d\theta : j \in R_k \right\} \\
 &= \arg \max_j \left\{ \int \sum_{r=1}^{m_j} g(\theta, U_j=r|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) J_{U_j=r}(\theta) d\theta : j \in R_k \right\},
 \end{aligned} \tag{A4}$$

where $g(\theta, U_j=r|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}})$ is the joint distribution of θ and U_j . This joint distribution is the product of the posterior for θ given the previous observations, $g(\theta|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}})$, and the probability of a response to item j given θ and the previous item responses, $f(U_j=r|\theta, u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}})$. From the local independence assumption that allows the IRT analysis, $f(U_j=r|\theta, u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) = f(U_j=r|\theta)$ since the items are independent given θ . Also from Equation 1, we have $J_{U_j}(\theta) \equiv E [J_{U_j}(\theta)] = \sum_{r=1}^{m_j} f(r|\theta) J_{U_j=r}(\theta)$.

So continuing from Equation A4,

$$\begin{aligned}
 \text{MEPWI}(\text{Observed}) &\equiv \arg \max_j \left\{ \int \sum_{r=1}^{m_j} g(\theta, U_j=r|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) J_{U_j=r}(\theta) d\theta : j \in R_k \right\} \\
 &= \arg \max_j \left\{ \int g(\theta|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) \sum_{r=1}^{m_j} f(U_j=r|\theta) J_{U_j=r}(\theta) d\theta : j \in R_k \right\} \\
 &= \arg \max_j \left\{ \int g(\theta|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) J_{U_j}(\theta) d\theta : j \in R_k \right\} \\
 &\equiv \text{MPWI}.
 \end{aligned} \tag{A5}$$

Notice, we proved that MEPWI and MPWI are equivalent selection methods. Therefore, as long as local independence holds, MEPWI is equivalent to MPWI.

Continuing from Equation A5, we add a new constant C^* that is independent of the maximization to create the complete Fisher's information. We also use

$g(\theta|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) = \sum_{r=1}^{m_j} g(\theta, U_j=r|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}})$ to show that $\text{MEPWI}(\text{Observed})$ is equivalent to $\text{MEPWI}(\text{Fisher})$:

$$\begin{aligned}
\text{MEPWI(Observed)} &\equiv \arg \max_j \left\{ \underbrace{\int g(\theta|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) I_{U_j}(\theta) d\theta}_{\text{MPWI}} : j \in R_k \right\} \\
&= \arg \max_j \left\{ \underbrace{\int g(\theta|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) I_{U_{i_1, U_{i_2}, \dots, U_{i_{k-1}}}}(\theta) d\theta}_{C^*} \right. \\
&\quad \left. + \int g(\theta|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) I_{U_j}(\theta) d\theta : j \in R_k \right\} \\
&= \arg \max_j \left\{ \int g(\theta|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) I_{U_{i_1, U_{i_2}, \dots, U_{i_{k-1}}, U_j}}(\theta) d\theta : j \in R_k \right\} \\
&= \arg \max_j \left\{ \int I_{U_{i_1, U_{i_2}, \dots, U_{i_{k-1}}, U_j}}(\theta) \sum_{r=1}^{m_j} g(\theta, U_j=r|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) d\theta : j \in R_k \right\} \\
&= \arg \max_j \left\{ \int I_{U_{i_1, U_{i_2}, \dots, U_{i_{k-1}}, U_j}}(\theta) \sum_{r=1}^{m_j} p(U_j=r|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) \right. \\
&\quad \left. \times g(\theta|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}, U_j=r) d\theta : j \in R_k \right\} \\
&= \arg \max_j \left\{ \sum_{r=1}^{m_j} p(U_j=r|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}) \int g(\theta|u_{i_1}, u_{i_2}, \dots, u_{i_{k-1}}, U_j=r) \right. \\
&\quad \left. \times I_{U_{i_1, U_{i_2}, \dots, U_{i_{k-1}}, U_j}}(\theta) d\theta : j \in R_k \right\} \\
&\equiv \text{MEPWI(Fisher)}. \tag{A6}
\end{aligned}$$

Therefore, under local independence, the MEPWI selection method using Fisher's information function is mathematically equivalent to MEPWI using the observed information function. Furthermore, $\text{MEPWI} \equiv \text{MPWI}$. Incidentally, it is worth noting that, if the constant C^* is not added, then the above proof shows that replacing the observed information function with Fisher's information function for just the next item is equivalent to MEPWI (Observed) as well.

References

- Andersen, EB. Discrete statistical models with social sciences applications. Amsterdam: North-Holland: 1980.
- Baker, BF.; Kim, S-H. Item response theory: Parameter estimation techniques. Vol. 2nd ed. New York: Marcel Dekker; 2004.
- Bock RD, Mislevy RJ. Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement* 1982;6:431–444.
- Bjorner JB, Chang C-H, Thissen D, Reeve BB. Developing tailored instruments: Item banking and computerized adaptive assessment. *Quality of Life Research* 2007;16:95–108. [PubMed: 17530450]
- Carlin, BP.; Louis, TA. Bayes and empirical Bayes methods for data analysis. Vol. 2nd ed. Boca Raton, FL: St. Lucie Press; 2000.
- Chang H-H, Ying Z. A global information approach to computerized adaptive testing. *Applied Psychological Measurement* 1996;20:213–229.
- Chen S-Y, Ankenmann RD, Chang H-H. A comparison of item selection rules at the early stages of computerized adaptive testing. *Applied Psychological Measurement* 2000;24:241–255.
- Choi SW. Firestar: Computerized adaptive testing (CAT) simulation program for polytomous IRT models. *Applied Psychological Measurement*. (in press).
- Donohue JR. An empirical examination of the IRT information of polytomously scored reading items under the generalized partial credit model. *Journal of Educational Measurement* 1994;31:295–311.
- Fan, M.; Hsu, Y. Utility of Fisher information, global information and different starting abilities in mini CAT; Paper presented at the annual meeting of the National Council on Measurement in Education; New York. 1996 Apr.

- Gelman, A.; Carlin, JB.; Stern, HS.; Rubin, DB. Bayesian data analysis. Vol. 2nd ed. Boca Raton, FL: Chapman & Hall/CRC Press; 2004.
- Lord, FM. Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum; 1980.
- Muraki E. Information functions of the generalized partial credit model. *Applied Psychological Measurement* 1993;17:351–363.
- Orlando M, Thissen D. Further investigation of the performance of S-X2: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement* 2003;27:289–298.
- Owen RJ. A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. *Journal of the American Statistical Association* 1975;70:351–356.
- Passos VL, Berger MPF, Tan FE. Test design optimization in CAT early stage with the nominal response model. *Applied Psychological Measurement* 2007;31:213–232.
- Penfield RD. Applying Bayesian item selection approaches to adaptive tests using polytomous items. *Applied Measurement in Education* 2006;19:1–20.
- Reeve BB. Special issues for building computerized-adaptive tests for measuring patient reported outcomes: The National Institute of Health's investment in new technology. *Medical Care* 2006;44:S198–S204. [PubMed: 17060828]
- Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph* 1969;17
- Thissen, D.; Mislevy, RJ. Testing algorithms. In: Wainer, H., editor. *Computerized adaptive testing: A primer*. Vol. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum; 2000. p. 101-133.
- van der Linden WJ. Bayesian item selection criteria for adaptive testing. *Psychometrika* 1998;2:201–216.
- van der Linden, WJ.; Pashley, PJ. Item selection and ability estimation in adaptive testing. In: vander Linder, WJ.; Glas, CAW., editors. *Computerized adaptive testing: Theory and practice*. Boston: Kluwer Academic; 2000. p. 1-25.
- van Rijn PW, Eggen TJHM, Hemker BT, Sanders PF. Evaluation of selection procedures for computerized adaptive testing with polytomous items. *Applied Psychological Measurement* 2002;26:393–411.
- Veerkamp WJJ, Berger MPF. Some new item selection criteria for adaptive testing. *Journal of Educational and Behavioral Statistics* 1997;22:203–226.
- Wainer H, Thissen D. Estimating ability with the wrong model. *Journal of Educational Statistics* 1987;12:339–368.
- Weiss DJ. Improving measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement* 1982;6:473–492.

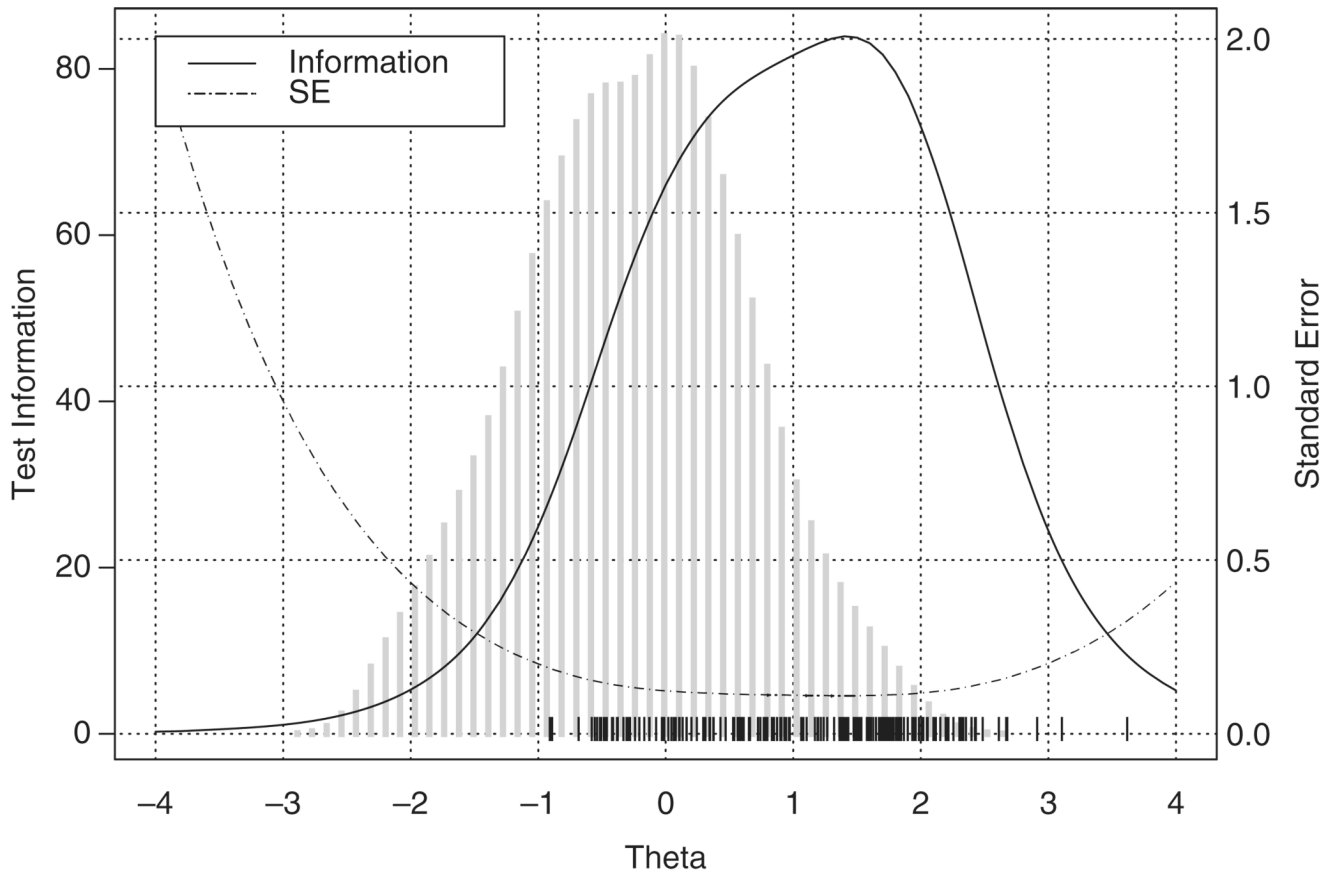


Figure 1.
Scale Information Function of Real Bank (62 Items) and Estimated Theta Distribution ($N = 730$)

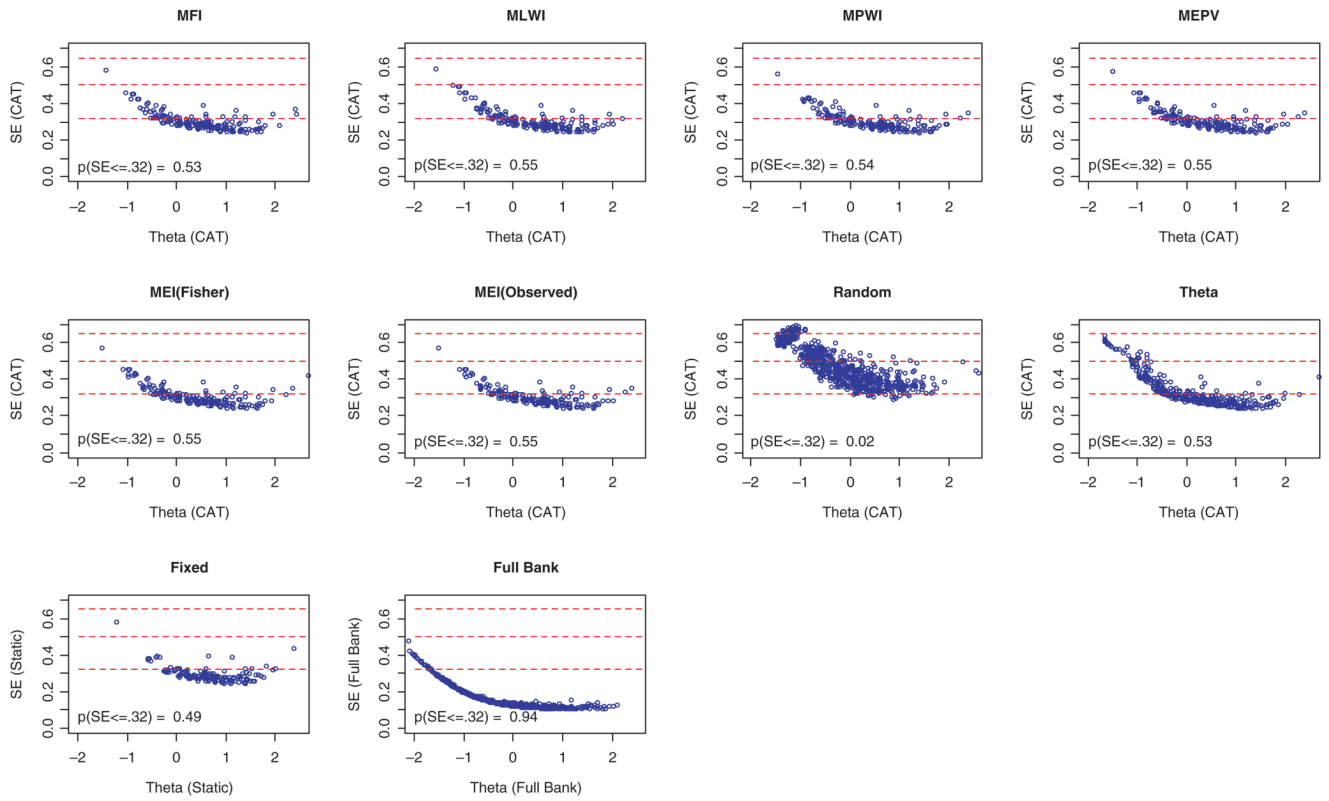


Figure 2.
SEM Calculated as Posterior Standard Deviation of Estimates: Real Bank, Real Data (5 Items)

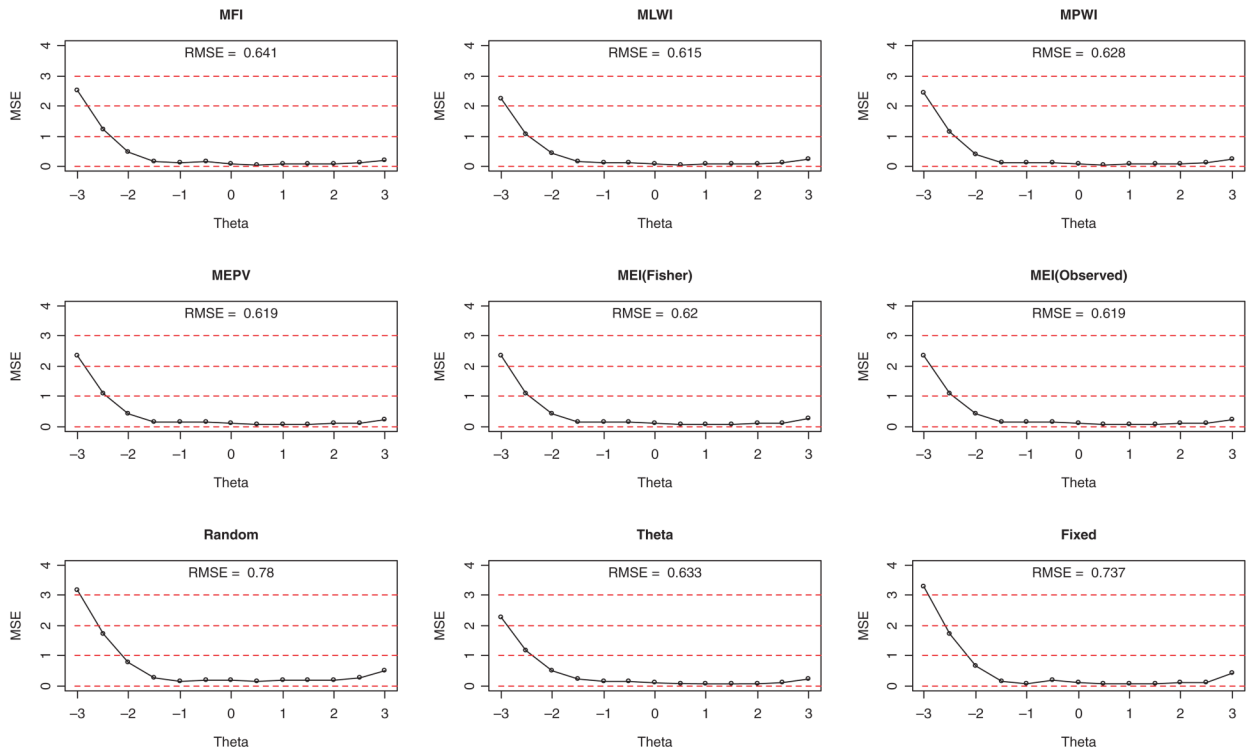


Figure 3. Mean Squared Error of Theta Estimates: Real Bank, Simulated Data (5 Items)

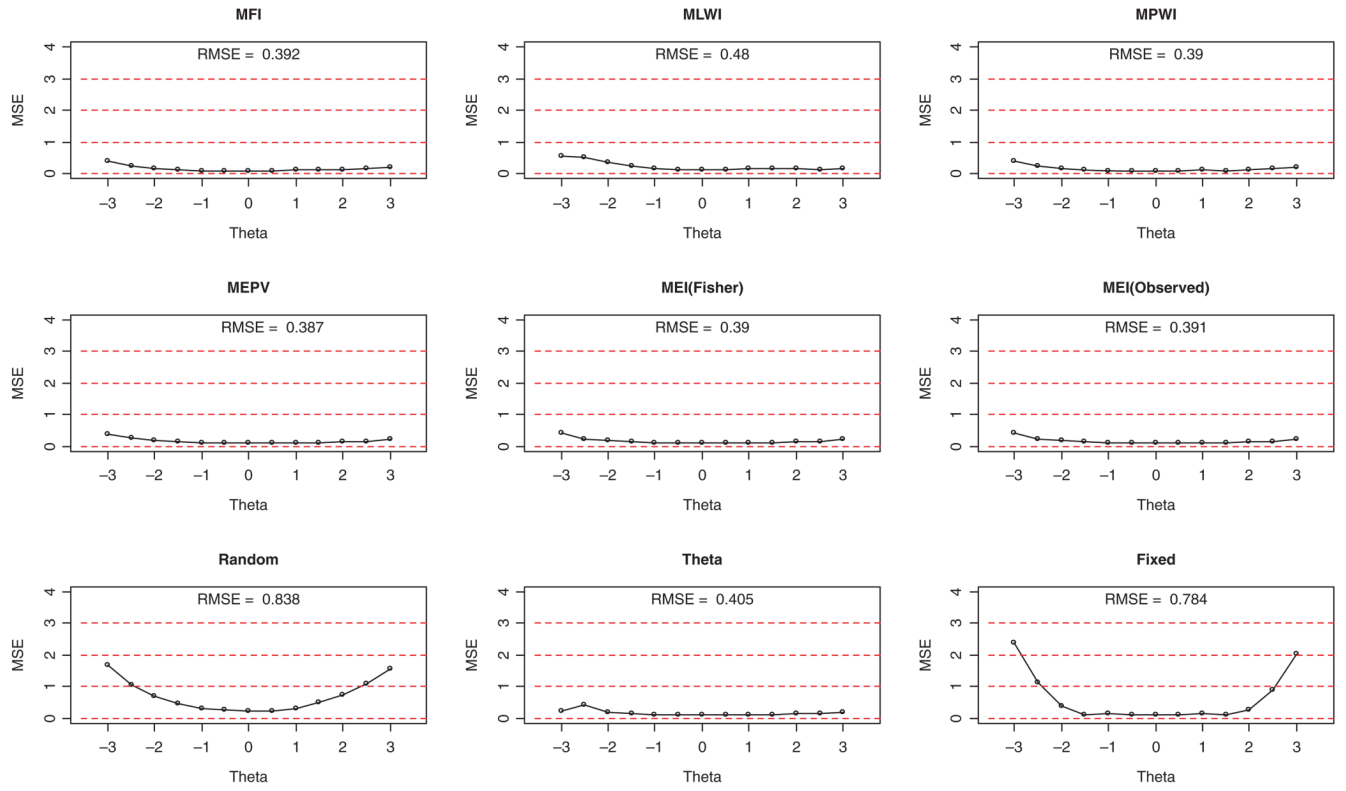


Figure 4.
Mean Squared Error of Theta Estimates: Simulated Peaked Bank, Simulated Data (5 Items)

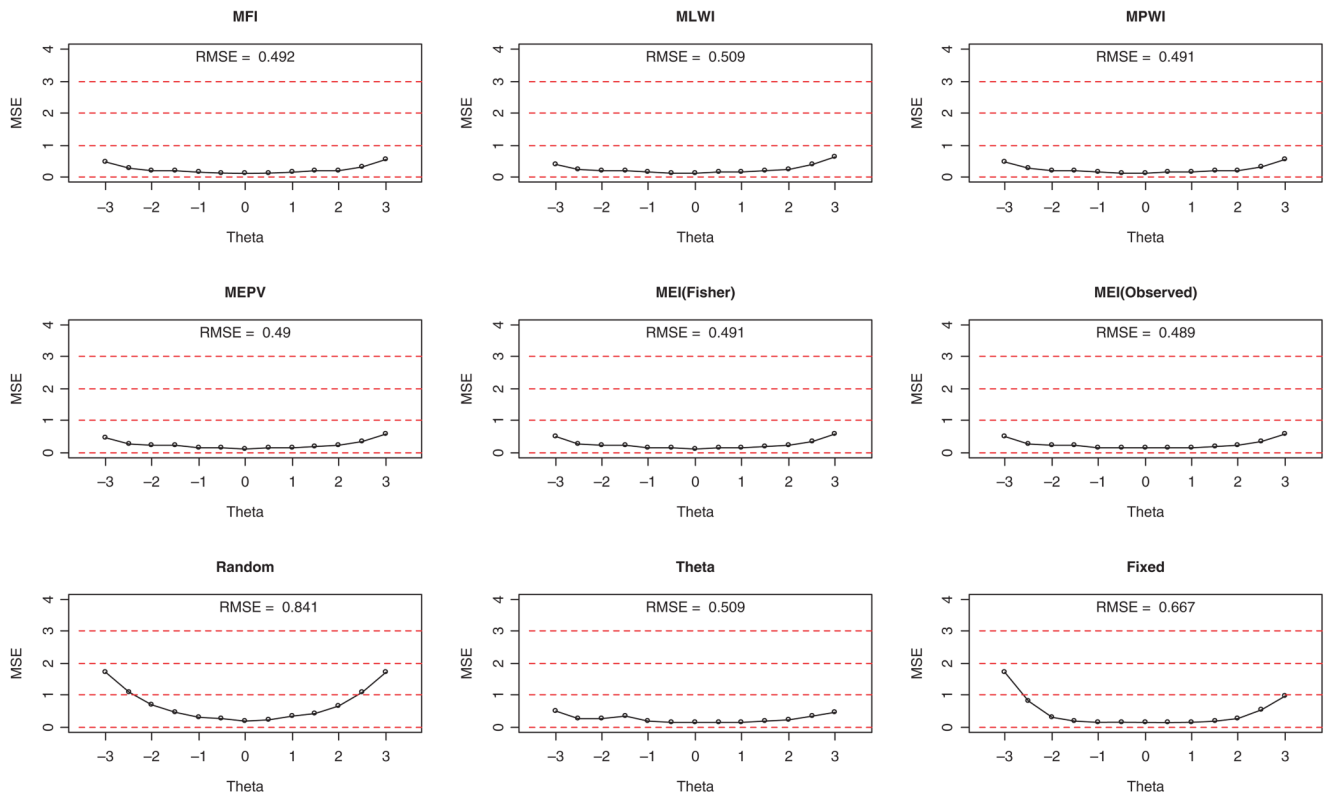


Figure 5. Mean Squared Error of Theta Estimates: Simulated Flat Bank, Simulated Data (5 Items)

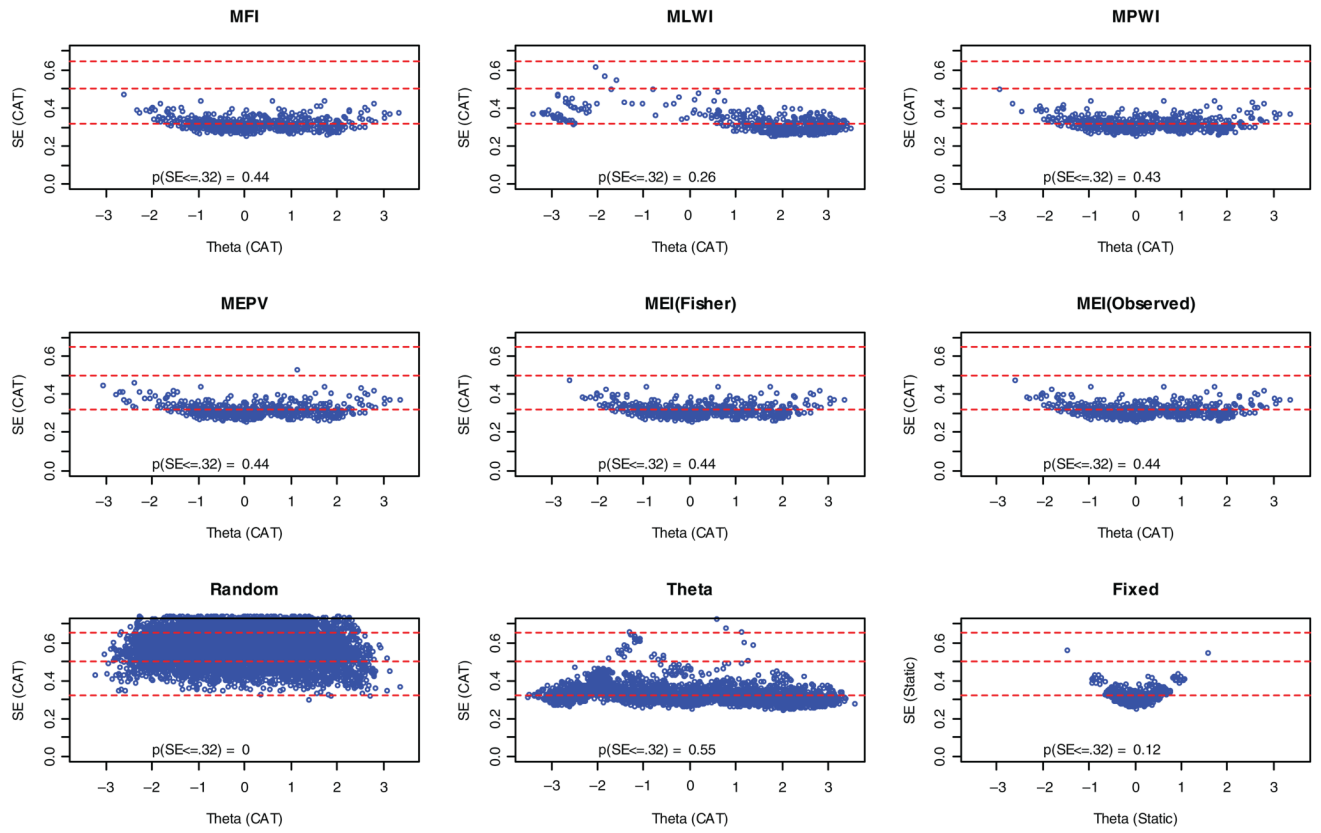


Figure 6. SEM Calculated as Posterior Standard Deviation of Estimates: Simulated Peaked Bank, Simulated Data (5 Items)

Table 1

Results for Real Bank, Real Data

Selection Criterion	5 Items			10 Items			20 Items		
	Mean SE^2	RMSD	CORR	Mean SE^2	RMSD	CORR	Mean SE^2	RMSD	CORR
MEI ^a	0.1463	0.3763	0.9069	0.0919	0.2745	0.9519	0.0578	0.1956	0.9770
MLWT ^b	0.1432	0.3736	0.9094	0.0911	0.2730	0.9530	0.0573	0.1955	0.9771
MPWT ^c	0.1396	0.3738	0.9080	0.0870	0.2627	0.9549	0.0563	0.1880	0.9783
MEPV ^d	0.1388	0.3598	0.9149	0.0881	0.2650	0.9546	0.0569	0.1963	0.9768
MEI (Fisher's) ^e	0.1388	0.3632	0.9134	0.0885	0.2677	0.9539	0.0569	0.1957	0.9769
MEI (Observed) ^f	0.1388	0.3616	0.9139	0.0878	0.2622	0.9555	0.0569	0.1951	0.9770
Random ^g	0.2369	0.4567	0.8565	0.1467	0.3337	0.9265	0.0842	0.2144	0.9700
Theta ^h	0.1461	0.3811	0.9048	0.0904	0.2745	0.9520	0.0561	0.1840	0.9791
Fixed ⁱ	0.1748	0.4281	0.8792	0.1213	0.3355	0.9267	0.0782	0.2522	0.9596

^aMaximum Fisher's information^bMaximum likelihood weighted information^cMaximum posterior weighted information^dMinimum expected posterior variance^eMaximum expected information using Fisher's information^fMaximum expected information using the observed information^gRandom selection^hSelected at the full bank estimated thetaⁱStatic form generated using expected information with respect to a standard normal distribution

Table 2

Results for Real Bank, Simulated Data

Selection Criterion	5 Items				10 Items				20 Items			
	BIAS	RMSE	CORR	BIAS	RMSE	CORR	BIAS	RMSE	CORR	BIAS	RMSE	CORR
MEI ^a	0.2000	0.6406	0.9626	0.1498	0.4898	0.9771	0.1102	0.3779	0.9854			
MLWT ^b	0.1806	0.6153	0.9658	0.1435	0.4803	0.9777	0.1074	0.3736	0.9857			
MPWT ^c	0.1860	0.6277	0.9645	0.1506	0.4899	0.9774	0.1089	0.3741	0.9858			
MEPV ^d	0.1849	0.6194	0.9651	0.1430	0.4782	0.9780	0.1079	0.3734	0.9857			
MEI (Fisher's) ^e	0.1825	0.6199	0.9653	0.1431	0.4787	0.9782	0.1077	0.3739	0.9857			
MEI (Observed) ^f	0.1847	0.6194	0.9650	0.1432	0.4777	0.9781	0.1081	0.3742	0.9857			
Random ^g	0.1984	0.7802	0.9444	0.1849	0.6141	0.9656	0.1451	0.4783	0.9780			
Theta ^h	0.1877	0.6325	0.9620	0.1415	0.4768	0.9777	0.1083	0.3745	0.9855			
Fixed ⁱ	0.2231	0.7372	0.9523	0.2076	0.6280	0.9641	0.1752	0.5127	0.9753			

^aMaximum Fisher's information

^bMaximum likelihood weighted information

^cMaximum posterior weighted information

^dMinimum expected posterior variance

^eMaximum expected information using Fisher's information

^fMaximum expected information using the observed information

^gRandom selection

^hSelected at known theta value

ⁱStatic form generated using expected information with respect to a standard normal distribution

Table 3

Results for Simulated Item Banks, Simulated Data

Bank	Selection Criterion	5 Items			10 Items			20 Items		
		BIAS	RMSE	CORR	BIAS	RMSE	CORR	BIAS	RMSE	CORR
PEAKED	MFI ^a	0.0283	0.3923	0.9822	0.0090	0.2744	0.9902	0.0032	0.2141	0.9938
	MLWI ^b	0.0678	0.4798	0.9724	0.0178	0.2916	0.9890	0.0051	0.2179	0.9936
	MPWI ^c	0.0261	0.3898	0.9822	0.0108	0.2747	0.9902	0.0033	0.2142	0.9938
	MEPV ^d	0.0232	0.3871	0.9822	0.0104	0.2741	0.9903	0.0033	0.2143	0.9938
	MEI (Fisher's) ^e	0.0299	0.3903	0.9824	0.0095	0.2743	0.9902	0.0032	0.2142	0.9938
FLAT	MEI (Observed) ^f	0.0283	0.3911	0.9823	0.0095	0.2745	0.9902	0.0032	0.2140	0.9938
	Random ^g	0.0095	0.8378	0.9233	0.0096	0.5554	0.9652	0.0000	0.3723	0.9834
	Theta ^h	0.0196	0.4050	0.9785	0.0060	0.2783	0.9898	0.0038	0.2156	0.9937
	Fixed ⁱ	0.0289	0.7840	0.9522	0.1060	0.5359	0.9746	0.0212	0.3403	0.9875
	MFI ^a	-0.0076	0.4921	0.9737	-0.0023	0.3628	0.9841	0.0007	0.2758	0.9902
FLAT	MLWI ^b	-0.0256	0.5093	0.9712	-0.0035	0.3641	0.9840	-0.0009	0.2771	0.9901
	MPWI ^c	-0.0086	0.4906	0.9737	0.0005	0.3620	0.9842	0.0014	0.2760	0.9902
	MEPV ^d	-0.0088	0.4905	0.9737	0.0007	0.3616	0.9842	0.0013	0.2757	0.9902
	MEI (Fisher's) ^e	-0.0093	0.4913	0.9738	-0.0022	0.3628	0.9841	0.0010	0.2760	0.9902
	MEI (Observed) ^f	-0.0082	0.4892	0.9739	-0.0019	0.3616	0.9842	0.0007	0.2758	0.9902
FLAT	Random ^g	0.0036	0.8410	0.9291	-0.0007	0.5867	0.9638	-0.0039	0.3986	0.9816
	Theta ^h	0.0059	0.5094	0.9695	0.0011	0.3644	0.9835	0.0002	0.2765	0.9901
	Fixed ⁱ	0.0470	0.6666	0.9606	0.0188	0.4596	0.9772	-0.0125	0.3329	0.9866

^aMaximum Fisher's information

^bMaximum likelihood weighted information

^cMaximum posterior weighted information

^dMinimum expected posterior variance

^eMaximum expected information using Fisher's information

^fMaximum expected information using the observed information

^gRandom selection

^hSelected at known theta value

ⁱStatic form generated using expected information with respect to a standard normal distribution