Item Response Theory Analyses of the Demonstrated Knowledge Items From the 1999 Medicare Current Beneficiary Survey

***Prepared for*:**

Sherry Terrell, Ph.D., Project Officer
Centers for Medicare and Medicaid Services
7500 Security Blvd.
M/S 3C-19-26
Baltimore, MD 21244-1850

Prepared by:

Carla Bann, Ph.D.
Research Triangle Institute
3040 Cornwallis Road
P.O. Box 12194
Research Triangle Park, NC 27709-2194

HCFA Contract No. 500-00-0024/002
RTI Project No. 07964.002

December 26, 2001

Item Response Theory Analyses of the Demonstrated Knowledge Items From the 1999 Medicare Current Beneficiary Survey

***Prepared for*:**

Sherry Terrell, Ph.D., Project Officer
Centers for Medicare and Medicaid Services
7500 Security Blvd.
M/S 3C-19-26
Baltimore, MD 21244-1850

Prepared by:

Carla Bann, Ph.D.
Research Triangle Institute
3040 Cornwallis Road
P.O. Box 12194
Research Triangle Park, NC 27709-2194

HCFA Contract No. 500-00-0024/002
RTI Project No. 07964.002

December 26, 2001

Table of Contents

Section **Page**

List of Tables

List of Figures

Executive Summary

The National Medicare Education Program (NMEP) seeks to inform Medicare beneficiaries about the Medicare program and related health plan options. An important aspect of assessing the effectiveness of the NMEP is to evaluate its impact on beneficiaries' knowledge of the plan choices. The Medicare Current Beneficiary Survey (MCBS) provides a potentially useful source of information about beneficiary knowledge. However, the knowledge items included in the MCBS have often been changed from year to year, making it difficult to assess possible improvements in knowledge as the result of programs, such as the NMEP. Item Response Theory (IRT), a statistical theory originally established in educational testing, may be used to create comparable knowledge scores over time, even when the same set of items has not been administered.

The goal of RTI's project, *Questionnaire Development and Cognitive Testing Using Item Response Theory*, is to develop a pool of knowledge items with established IRT parameters. Developing this item pool will allow the knowledge items in the MCBS to be rotated from year to year while still providing comparable scores across years. This approach would allow the MCBS knowledge index to remain adaptable to changes in the Medicare program while consistently monitoring changes in knowledge over time.

The purpose of this report is to evaluate the knowledge items administered during the 1999 MCBS for possible inclusion in the item pool. We conducted IRT analyses to evaluate the psychometric properties of these items. Because conventional IRT models assume that the scale being analyzed is unidimensional, factor analyses were conducted first to evaluate the dimensionality of the knowledge items. The factor analysis results indicated that items from the Beneficiary Knowledge (BK) Round-26 and Beneficiary Needs (BN) Round-27 supplements formed two separate factors and therefore should be analyzed separately. With only three items, IRT can produce unstable estimates, so the 3-item BN knowledge quiz could not be analyzed using IRT. Therefore, the remaining analyses in the report utilized only the 7-item BK knowledge quiz.

Overall, the IRT analyses suggested that the BK demonstrated knowledge items are good candidates for inclusion in the item pool. The IRT item parameters indicated that all seven items showed good discrimination ability and therefore were related to the underlying construct. The knowledge items had a variety of difficulty parameters, ranging from –1.27 to 0.46. These results suggested that the quiz cannot effectively discriminate between beneficiaries with higher knowledge levels ($\theta > 0.46$). Adding some more difficult items would improve the ability of the quiz to effectively discriminate knowledge for a wider range of beneficiaries.

In addition to evaluating the item parameters, IRT-based scores were computed for each respondent. Because both the item parameters and the individual responses are included in the calculation of the score, the IRT-based scores are more precise estimates of knowledge than Classical Test Theory scores.

Differential item functioning analyses were also conducted to compare the functioning of the items for three groups of beneficiaries: (1) male vs. female, (2) aged vs. disabled, and (3) those with none vs. some managed care enrollment. The results indicate that five of the seven items were differentially difficult according to whether the respondent was enrolled in managed care during the past year.

A limitation of the current analyses is that the 1999 MCBS the quiz items were administered only to beneficiaries who were in their first year of participation in the survey. Therefore, the IRT parameters may only be appropriate for individuals who have just begun the study. To determine whether these IRT parameters are applicable to beneficiaries in all four years, similar analyses could be conducted with data from the BK supplement included in the 2000 MCBS that administered the seven quiz items to all MCBS participants.

In summary, the analyses in this report have identified and evaluated a set of potentially useful items for inclusion in the item pool. However, creating a large pool of items would require the development and testing of additional items. For example, while the 7-item quiz can assess knowledge of some aspects of managed care, Medicare beneficiaries may need other types of knowledge to effectively navigate the Medicare system. Therefore, additional items are required to measure other aspects of Medicare-related knowledge.

In addition to evaluating the existing MCBS knowledge items, another goal of the project was to develop a set of new knowledge items. Several new knowledge items were created and cognitively tested, addressing topics such as eligibility for and structure of Original Medicare, beneficiary rights and protections, and how to get more information and assistance (Uhrig et al., 2001). To capitalize on the benefits of IRT, we recommend that the next step in the development of the MCBS knowledge item pool be to conduct a pre-test in which all of the newly developed items are administered to a large sample of respondents. The respondents selected for the pre-test should be representative of the population that will eventually be administered the MCBS items. Once the data have been collected, the items could be calibrated and used to develop a set of equivalent forms that would allow different sets of respondents to receive different knowledge questions, while still receiving comparable scores. Calibration of the items would also make it possible to change the items from year to year and potentially to intersperse new items during future years.

1.0 Background

The National Medicare Education Program (NMEP) seeks to educate beneficiaries about the Medicare program and related health plan choices. In 1997, the Centers for Medicare and Medicaid Services (CMS) developed a new education campaign to inform beneficiaries about the choices available to them and to make them aware of options for finding more information, such as 1-800-Medicare or the Medicare website. To evaluate the effectiveness of the NMEP measures are needed to assess Medicare beneficiaries' knowledge of the available health plan choices.

With its large, longitudinal sample of Medicare beneficiaries, the Medicare Current Beneficiary Survey (MCBS) provides a potential source of data for measuring beneficiary knowledge and thereby evaluating the impact of programs, such as the NMEP, in changing beneficiary knowledge. However, traditionally, knowledge questions on the MCBS have changed from year to year to address the newest features of Medicare health plans and changes in Medicare benefits, and to adapt to the changing priorities and goals of CMS. The changing content makes it difficult to measure improvement or decline in beneficiary knowledge from year to year and therefore to evaluate the effectiveness of interventions designed to increase beneficiary knowledge.

Item Response Theory (IRT) may be used to remedy this problem by assigning a comparable metric to knowledge measures that differ from year to year in future rounds of the MCBS. The use of IRT to develop and evaluate measures has been well established in the field of educational testing. For example, many large-scale testing programs, such as the Law School Admission Test (LSAT) and Graduate Record Examination (GRE), use IRT to equate different test forms. With its roots in educational testing, IRT is particularly well suited to this application. As with traditional educational tests, the goal of this project is to measure knowledge of a particular topic.

The goal of RTI's project, *Questionnaire Development and Cognitive Testing Using Item Response Theory (IRT)*, is to develop a pool of knowledge items with known IRT item parameters. Developing this item pool will allow the knowledge items in the MCBS to be rotated from year to year while still providing comparable scores across years. Further, it allows for the addition of new items once the IRT parameters of the current items have been established. This approach allows the MCBS knowledge index to remain adaptable to changes in the Medicare program while also allowing for the analysis of change in knowledge over time.

The 1999 MCBS contained knowledge items in the Round-26 Supplement that are potential candidates for inclusion in the final item pool. IRT analyses were conducted to evaluate the psychometric properties of these items. The results of these analyses are outlined in this report and recommendations are provided concerning the usability of the items for future rounds of the MCBS.

2.0 Introduction to Item Response Theory (IRT)

This section provides a brief introduction to Item Response Theory (IRT; Embretson & Reise, 2000; Hambleton, Swaminathan, & Rogers, 1991) to serve as a basis for understanding the analyses described in this report. IRT assumes that a test or questionnaire is measuring a single underlying construct (denoted as θ). IRT uses responses to questionnaire items to estimate an individual's level on the underlying construct. For example, in the present study, responses to the MCBS demonstrated knowledge items may be used to estimate a beneficiary's knowledge of the Medicare program.

IRT uses a model to describe the relationship between an individual's response to an item and the underlying construct. There are several different types of models used in IRT. Because the items in this study were coded into two categories (correct vs. incorrect), the three IRT models for dichotomous items were used for the analyses described in this report. For each model, a set of parameters is estimated to describe the characteristics of each item. The first model is the **3-parameter logistic (3PL) model**. As the name implies, the 3-PL model estimates three item parameters for each item. The first parameter is the **threshold** or *b* parameter. This parameter represents the difficulty of the item; items with higher *b* parameters are more difficult. The next parameter is the **slope** or *a* parameter. The slope quantifies how related the item is to the construct being measured by the scale. The last parameter is the **guessing** or *c* parameter. This parameter is used to try to explain why some individuals with low ability answer an item correctly.

Another model for dichotomous items is the **2-parameter logistic (2PL) model**. The 2PL model is identical to the 3PL model except that only the threshold (*b*) and slope (*a*) parameters are estimated for each item. The guessing (*c*) parameter is assumed to be equal to 0. The **1-parameter logistic (1PL) model** adds an additional restriction to the 2PL model. In the 1PL model only the threshold (*b*) parameter is estimated separately for each item. The slope (*a*) parameter is assumed to be equal for all items and the guessing (*c*) parameter is set equal to 0.

IRT and Classical Test Theory utilize two different approaches for computing scores. Traditionally, in Classical Test Theory, only the individual's responses to the items are used in the calculation of the score. For example, a summed score is computed by totaling the number of items an individual answered correctly. In contrast, IRT uses information about both the individual's responses and the item parameters to compute the score. For example, an individual answering two items with higher slopes correctly would receive a higher score than someone who answered two items with lower slopes correctly.

One type of IRT analysis, **differential item functioning** (*dif*) analysis, can be used to identify items that function differently for different groups. Investigating possible *dif* is an important step in evaluating the psychometric properties of a scale because the presence of *dif* reduces the validity of a scale. Validity may be defined as "how well (the scale) measures what it purports to measure" (Nunnally & Bernstein,

1994). If the items on a scale exhibit *dif*, they are no longer measuring only the construct of interest, but also at least one additional factor related to group membership.

Dif analysis is especially valuable when one wishes to compare or combine results across various populations. For example, if particular items are more difficult for one group than another, the first group may have lower scores simply because of the particular items that were included on the scale. Their lower scores may not necessarily indicate that they have less knowledge.

Item response theory analyses may be used to uncover two types of differential item functioning. First of all, *dif* with respect to the threshold (*b*) parameters of an item (**threshold-related *dif***) indicates that the two groups differ in how difficult the item is for them. The presence of this type of *dif* suggests that the results for the scale should simply be reported separately for the two groups. However, the other type of *dif,* **slope-related *dif*,** indicates that the item is differentially related to the underlying construct for the two groups. Severe slope-related *dif* may suggest a different factor structure for the two groups. Slope-related *dif* is more detrimental than threshold-related *dif* and indicates that the item should not be used to compare the two groups.

In this report, IRT analyses were conducted to evaluate the psychometric properties of the knowledge items included in the 1999 Medicare Current Beneficiary Survey (MCBS).[1] The IRT analysis results were then used to calculate IRT-based scores for the respondents. In addition, the possibility of differential item functioning was investigated. The next section describes the knowledge items included in these analyses.

[1] Classical test theory analyses (e.g., percentage correct, item-total correlations) of these items will be described in a separate report.

3.0 Knowledge Items Included in the 1999 MCBS

The 1999 MCBS contains two types of items that may be used as measures of knowledge. The first set of knowledge measures asks beneficiaries to indicate how much they *feel* they know about a particular topic. The beneficiaries respond to the questions by rating their knowledge using a 5-point scale, ranging from "Just about everything I need to know" to "Almost none of what I need to know." These constitute self-reported measures of knowledge. The other knowledge measures have been called demonstrated knowledge items because they require beneficiaries to verify their knowledge by providing the correct answer to question that has a single correct answer.

This report focuses on the demonstrated knowledge items only. These items require beneficiaries to demonstrate that they do know a particular piece of information, rather than simply providing their perceptions of the adequacy of their knowledge. For this reason, demonstrated knowledge items may be considered a more precise measure of knowledge. These items have one correct answer and therefore, individual responses can be scored as correct or incorrect.

Table 1 provides a list of the seven demonstrated knowledge items that were administered in the Beneficiary Knowledge (BK) Supplement during Round 26 of the MCBS. The item numbers displayed are those used in the BK supplement[2]. In the Beneficiary Information Needs (BN) supplement during Round 27 of the MCBS, three additional demonstrated knowledge items were administered (see Table 2). Again, the item numbers are those from the BN supplement[3]. During the 1999 MCBS, only respondents who were in their first year of the panel survey received the demonstrated knowledge items in the BK and BN supplements.

All of the demonstrated knowledge items included in Tables 1 and 2 are true/false questions; the correct answers are enclosed in parentheses. For each question, respondents have the option of indicating that they don't know the answer. For the analyses in this report, each of the items was recoded as correct or incorrect with don't know responses considered to be incorrect. This approach is consistent with previous scoring algorithms used for this type of questions (Hibbard, Jewett, Englemann, & Tusler, 1998; Bann, Lissy, Keller, Garfinkel, & Bonito, 2000; McCormack, Anderson, Daugherty, Ross, Kuo, & Garfinkel, 2000).

Before the IRT analyses could be conducted, it was important to verify that these items met the IRT assumption of unidimensionality. Therefore, we conducted factor analyses to explore the dimensionality of these items.

[2] Item BK 45 was not included in the 1999 MCBS.

[3] Item BN 17 concerned coverage of mammograms and was administered only to women. Because this item was not administered to all respondents, it was omitted from these analyses.

Table 1. List of Demonstrated Knowledge Items Included in the 1999 Beneficiary Knowledge (BK) Supplement (Round 26).

Item #	Short Description	Item
BK43	Plan options available	Most people covered by Medicare can select among different kinds of health plan options within Medicare. *(True)*
BK44	Medicare alone pays all expenses	Medicare without a supplemental insurance policy pays for all of your health care expenses. *(False)*
BK46	Medicare is offering more information	The Medicare program has begun to offer more information and help in order to answer your Medicare questions. *(True)*
BK47	Can report complaints about HMOs	People can report complaints to Medicare about their Medicare managed care plans (HMOs) or supplemental plans if they are not satisfied with them. *(True)*
BK48	Limited choice of doctors in HMOs	If someone joins a Medicare managed care plan (HMO) that covers people on Medicare they have limited choices about which doctors they can see. *(True)*
BK49	Can drop HMO and still be covered	If someone joins a Medicare managed care plan (HMO) that covers people on Medicare, they can change or drop the plan and still be covered by Medicare. *(True)*
BK50	HMOs cover more services	Medicare managed care plans (HMOs) that cover people on Medicare often cover more health services, like prescribed medicines, than Medicare without a supplemental policy. *(True)*

Table 2. List of Demonstrated Knowledge Items Included in the 1999 Beneficiary Information Needs (BN) Supplement (Round 27).

Item #	Short Description	Item
BN16	Colorectal cancer screening	Medicare covers colorectal cancer screening. *(True)*
BN18	Managed care plans	Medigap or supplemental insurance is the same as a Medicare managed care plan. *(False)*
BN19	Flu shot	Medicare covers an annual flu shot. *(True)*

4.0 Factor Analyses

We conducted a factor analysis to explore whether all 10 items from the two supplements (seven from BK and three from BN) could be combined into a single scale. If all 10 items comprise a unidimensional scale, then it may be reasonable to analyze them all together using IRT.

Because the items were coded dichotomously (i.e., correct vs. incorrect), tetrachoric correlations were computed to compare the relationships of the items. The correlations are presented in Table 3; values greater than or equal to 0.30 are shown bolded. Based on the correlations, Items 48, 49, and 50 appear to be the most highly related of the 10 items ($r \geq 0.60$). Each of these items addresses properties of managed care plans. Examining the correlation matrix also suggests that, with few exceptions, the BK items are more highly related to each other than to the BN items. The same pattern exists for the BN items.

Next, these tetrachoric correlations were used as the input data for a maximum likelihood factor analysis using promax rotation. The factor analysis results suggest that the items cluster into two separate factors. Factor loadings for the two factors are presented in Table 4; loadings greater than 0.30 are shown bolded. The first factor contains all seven of the items from the BK supplement and the second factor contains the three items from the BN supplement. The correlation between the factors was 0.46, indicating that they are moderately related. These results suggest that it may be inappropriate to combine the items from the two supplements into a single scale.

Another concern about combining the 10 items into one scale is the time lag between data collection for the two supplements. The period of data collection for the BK supplement was January through April 2000 while the BN supplement was collected from May through August 2000. Given these dates of collection, for some respondents, there could potentially have been a seven-month time lag between the administration of the two sets of items. During this time period, the respondents could have obtained more knowledge of the Medicare program. Therefore, an item in the BN supplement could appear to be easier than it would have if it had been included in the earlier BK supplement.

Overall, given the factor analysis results and the time lag between administration of the items, we decided that the two sets of items should be analyzed separately. However, IRT can produce unstable estimates when only three items are analyzed. Therefore, for this report, only the seven demonstrated knowledge items from the BK supplement were analyzed using IRT.

Table 3. Tetrachoric Correlations of the MCBS BN[1] (Round 27) and BK[2] (Round 26) Demonstrated Knowledge Questions.

Item	BN16	BN18	BN19	BK43	BK44	BK46	BK47	BK48	BK49	BK50
BN16. Colorectal cancer screening	1.00									
BN18. Managed care plans	**0.37**	1.00								
BN19. Flu shot	**0.47**	**0.34**	1.00							
BK43. Plan options available	0.27	0.23	0.24	1.00						
BK44. Medicare alone pays all expenses	0.26	**0.36**	**0.31**	**0.40**	1.00					
BK46. Medicare is offering more information	0.26	0.20	0.24	**0.42**	**0.41**	1.00				
BK47. Can report complaints about HMOs	0.18	0.21	0.25	**0.41**	**0.40**	**0.46**	1.00			
BK48. Limited choice of doctors in HMOs	0.21	**0.35**	0.22	**0.35**	**0.54**	**0.31**	**0.47**	1.00		
BK49. Can drop HMO and still be covered	0.25	**0.32**	0.23	**0.43**	**0.44**	**0.37**	**0.44**	**0.64**	1.00	
BK50. HMOs cover more services	0.20	**0.30**	0.20	**0.40**	**0.37**	0.29	**0.40**	**0.60**	**0.68**	1.00

[1] Beneficiary Needs
[2] Beneficiary Knowledge

SOURCE: Centers for Medicare and Medicaid Services, Medicare Current Beneficiary Survey 1999 Access to Care and Supplemental Files.

Table 4. Factor Loadings for the MCBS BN (Round 27) and BK (Round 26) Demonstrated Knowledge Items.

Item	Factor 1: BK Items	Factor 2: BN Items
BN16. Colorectal cancer screening	-0.01	**0.65**
BN18. Managed care plans	0.22	**0.41**
BN19. Flu shot	-0.02	**0.68**
BK43. Plan options available	**0.43**	0.23
BK44. Medicare alone pays all expenses	**0.44**	0.29
BK46. Medicare is offering more info	**0.34**	0.28
BK47. Can report complaints about HMOs	**0.49**	0.16
BK48. Limited choice of doctors in HMOs	**0.77**	0.00
BK49. Can drop HMO and still be covered	**0.85**	-0.04
BK50. HMOs cover more services	**0.81**	-0.08

Before beginning the IRT analyses, we conducted a second factor analysis to verify that the seven BK items were in fact unidimensional. The factor analysis results indicated that the seven items do comprise one factor; the factor loadings are presented in Table 5. All items have loadings of 0.50 or greater. Item 49 (Can drop HMO and still be covered) has the highest factor loading (loading=0.81), indicating that it is most related to the underlying construct. Four of the seven items (Items 47-50) address topics related to managed care plans, suggesting that overall the factor seems to primarily represent knowledge of managed care plans.

Table 5. Factor Loadings for the MCBS BK (Round 26) Knowledge Items Only.

Item	Factor 1: Managed Care
BK43. Plan options available	**0.55**
BK44. Medicare alone pays all expenses	**0.61**
BK46. Medicare is offering more information	**0.50**
BK47. Can report complaints about HMOs	**0.60**
BK48. Limited choice of doctors in HMOs	**0.77**
BK49. Can drop HMO and still be covered	**0.81**
BK50. HMOs cover more services	**0.75**

The factor analysis results indicated that the seven knowledge items formed a single factor, thereby providing evidence that the knowledge scale meets the IRT assumption of unidimensionality[4]. Therefore, the next step was to conduct IRT analyses to investigate the psychometric properties of the items.

[4] Multidimensional IRT models do exist, however, many of the techniques for estimating these models are still under development and have not been implemented in commercial software.

5.0 Item Response Theory Analyses

The first step in the IRT analyses was to identify the best-fitting model. Because the knowledge items are dichotomous, these three models were fit to the data: (a) 1-parameter logistic model, (b) 2-parameter logistic model, and (c) 3-parameter logistic model. The IRT models were computed using Multilog Version 6.0 which utilizes the maximum marginal likelihood estimation technique for estimating the item parameters (Thissen, 1991).

To determine the most appropriate IRT model, the statistical fit of competing models was compared. Specifically, the difference between the negative twice the log-likelihood values for two alternative models was interpreted as a chi-square with degrees of freedom equal to the number of additional parameters estimated by the less-constrained model. In other words, the test for the comparison between the 1PL and 2PL models may be expressed as:

$$\chi^2 = (\text{-2*log-likelihood}_{1PL}) - (\text{-2*log-likelihood}_{2PL}) = 844.0 - 616.1 = 227.9$$

This result indicates that the 2PL model provides a significantly better fit than the 1PL model ($\chi^2(6) = 227.9$, $p < .001$). Next, the 2PL model was compared to the 3PL model; the results indicate that the 3PL model does not provide a significant improvement in fit over the 2PL model ($\chi^2(7) = 0$, $p > .05$). In fact, the values for negative twice the log-likelihood were identical in the two models. When the c parameters were freely estimated in the 3PL model, they remained equal to 0 as in the 2PL model. Overall, the model comparisons indicated that the 2PL model provided the best fit and therefore this model was used to evaluate the knowledge items.

The item parameters estimated using the 2PL model are presented in Table 6. As mentioned earlier, the slope or *a* parameter indicates the item's discrimination while the threshold or *b* parameter measures the item's difficulty level. As shown in Table 6, all seven items demonstrated good discrimination with *a* parameters close to or greater than 1.0. Of the seven items, Item 46 (Medicare is offering more information) had the smallest slope parameter ($a = 0.99$), indicating that this item is the least related to the underlying construct. This finding is consistent with the factor analysis results earlier in which Item 46 had the smallest factor loading. Item 49 (Can drop HMO and still be covered) obtained the largest slope parameter ($a = 2.56$) and again this is consistent with the factor analysis results in which this item also had the largest factor loading.

Comparing the *b* parameters is useful for evaluating the relative difficulties of the items. Item 44 (Medicare alone pays all expenses) was by far the easiest item with a difficulty parameter equal to –1.27, suggesting that perhaps this is a basic concept that most beneficiaries understand. The most difficult item was Item 50 (HMOs cover more services). This item had a difficulty parameter equal to 0.46, indicating that it is only marginally difficult. This item may be more difficult than the other items because in order to answer it correctly beneficiaries must have knowledge of both Original Medicare and managed care plans.

Table 6. IRT Parameters (and Standard Errors) for the MCBS 1999 Supplemental BK (Round-26) Items Estimated Using the 2PL Model.

Item	Parameters	
	Slope (*a*)	**Threshold (*b*)**
BK43. Plan options available	1.14 (0.06)	0.26 (0.04)
BK44. Medicare alone pays all expenses	1.33 (0.06)	-1.27 (0.06)
BK46. Medicare is offering more information	0.99 (0.06)	-0.08 (0.05)
BK47. Can report complaints about HMOs	1.23 (0.06)	-0.26 (0.04)
BK48. Limited choice of doctors in HMOs	1.89 (0.08)	-0.31 (0.03)
BK49. Can drop HMO and still be covered	2.56 (0.10)	0.16 (0.02)
BK50. HMOs cover more services	2.14 (0.09)	0.46 (0.03)

The information provided by the item parameters may also be represented graphically. For example, the item parameters were used to create **item characteristic curves** for each item (see Figures A-1 to A-7 in Appendix A). The X-axis of an item characteristic curve represents the different levels of the construct and is denoted as theta (θ). For the seven-item BK quiz, theta represents knowledge of the Medicare program. The Y-axis of the item characteristic curve represents the probability of answering the item correctly. The curve shows the relationship between knowledge (theta) levels and the probability of answering the item correctly. For each of the item characteristic curves, as theta increases, thereby indicating higher knowledge levels, the probability of answering the item correctly increases. In other words, beneficiaries with more knowledge are more likely to answer the item correctly.

The values of the item parameters shown in Table 6 determine the shape and location of the curves. The item's difficulty or *b* parameter determines the location of the curve along the X-axis. This parameter indicates the value of theta at which individuals have a 50% probability of giving a correct response. To illustrate this point, dashed lines were included in Figures A-1 to A-7 to indicate the location of the b parameter for each item. For example, by examining the item characteristic curve for Item 46, we can see that individuals with a theta of just below 0 have a 50% probability of answering this item correctly (see Figure A-3). The value of theta at this point is equal to the item's difficulty parameter. (The difficulty parameter for Item 46 is actually equal to –0.08).

Items with larger *b* parameters will be shifted to the right on the X-axis, indicating that an individual needs to have a higher knowledge (theta) level to have a 50% probability of answering the item correctly. On the other hand, items with smaller *b*

parameters will be shifted to the left side of the X-axis because a lower knowledge level is required to still have a 50% probability of giving a correct response.

The item's *a* parameter determines the slope or steepness of the curve. This parameter indicates the ability of the item to differentiate between individuals with knowledge levels above or below the point where theta is equal to *b*. Items with smaller slopes tend to be flatter while items with larger slopes are steeper.

For comparison purposes, the item characteristic curves for all seven items are presented in Figure A-8. It appears that the curves for Items 48, 49, and 50 are fairly steep. As shown in Table 6, these three items had the highest slope or *a* parameters. The curves for Items 43, 44, 46, and 47 are somewhat flatter due to their smaller *a* parameters. Therefore, these four items do not make as sharp a distinction between those who are above or below their respective *b* values.

It is also apparent from the graphs that Item 44 is the easiest item; respondents with a knowledge level of only –1.27 or above have at least a 50% probability of getting this item correct. In contrast, Item 50 is the most difficult item, requiring respondents to have a knowledge level of at least 0.46 to obtain a 50% chance of answering the item correctly. Viewing all seven curves at once also reveals that the item difficulties are clustered at the lower half of the knowledge scale, indicating that the items have only mid to low levels of difficulty. None of the items have difficulties greater than 0.46. Therefore, the scale cannot effectively discriminate among respondents with higher levels of knowledge. It may be useful to add some more difficult items to the scale to improve its ability to measure knowledge for a wider range of beneficiaries.

The next type of curve that may be used to investigate the properties of the items is an **item information curve**. Information is a statistic in IRT that assesses the amount of psychometric information the item (or scale) provides for respondents at different levels on the underlying construct. One common use of information is in computerized-adaptive testing where the test is tailored to individual respondents by selecting items that are most informative at the respondent's ability level.

The information curves for all seven items are shown in Figure A-9. Items are most informative at theta values close to their *b* parameters. For example, the information curve for Item 50 peaks at theta close to 0.5. Therefore, this item is most useful for measuring knowledge of respondents with slightly above average knowledge levels.

Information is also a function of the item's discrimination ability; items with higher slope or *a* parameters are more informative. For example, as shown in Figure A-9, Item 49 has the highest peak and therefore provides the greatest amount of information. As shown in Table 6, this item also has the largest slope parameter ($a = 2.56$).

Another useful graph is the **test information curve**. This curve combines the information curves for all of the items and indicates at which ability levels the entire test

is most informative. Examining the test information curve for the MCBS knowledge index shown in Figure 3 reveals that this index provides the most information for respondents with average knowledge levels (thetas ranging from –0.5 to 0.5).

Information is inversely related to standard error of measurement; as information decreases, the standard error of measurement increases. Figure A-11 displays the **test standard error of measurement curve**. As shown in this curve, the standard error of measurement for the entire knowledge index is lowest at theta close to 0, indicating that the index is most precise for individuals at this knowledge level.

5.1 IRT-Based Scores

If desired, the results from the IRT analyses may be used to assign scores to the respondents. Table B-1 in Appendix B contains a list of the IRT-based knowledge scores calculated using the 2PL model. IRT scores are assigned according to the individual's pattern of responses to the items. The first column in Table B-1 lists the set of observed response patterns to the seven knowledge items. For each of the response patterns, a 0 indicates that the item was answered incorrectly and a 1 indicates that the item was answered correctly. For example, a response pattern of 0010100 indicates that the 3rd and 5th items (Items 43 and 48) were correct and the rest of the items were incorrect.

As mentioned in Chapter 2, IRT and Classical Test Theory utilize two different approaches for computing scores. IRT uses information about both the individual's responses and the item parameters to compute the score, while Classical Test Theory only considers the individual's responses to the items. In the example above, using Classical Test Theory, the respondent with a response pattern of 0010100 would receive a summed score of 2 because he or she answered two items correctly. However, in the current analysis, this individual would receive an IRT-based score of –0.63 (see Table B-1).

Another respondent who also responded correctly to two items may receive a different IRT-based score, depending on which two items were answered correctly. For example, an individual who answered Items 49 and 50 correctly (i.e., had a response pattern of 0000011) would receive a score of –0.26 instead. The difference in the scores for these two individuals is the result of the inclusion of the item parameters in the score calculation. Specifically, an individual who correctly answered two more discriminating items would be assigned a higher score than an individual who correctly answered two less discriminating items. In IRT-based scoring, items with higher slope (a) parameters are given more weight in the score computation.

6.0 Differential Item Functioning Analyses

This section summarizes the results from IRT *dif* analyses conducted on responses to the seven BK demonstrated knowledge items. First, analyses were conducted to investigate possible differential item functioning (*dif*) between males and females. The next set of analyses explores potential *dif* among two groups: (1) respondents who are eligible for Medicare because of their age and (2) respondents who are eligible for Medicare because they are disabled or have end stage renal disease. The final set of *dif* analyses compares respondents who were enrolled in a managed care plan during the past year with respondents who had no managed care enrollment during that time.

6.1 Gender

Item response theory (IRT) analyses were conducted in an effort to identify any of the knowledge items that functioned differently for males and females. An identical procedure was followed for each item. First, IRT parameter estimates were obtained for a 2PL model in which all parameter estimates were constrained to be equal across the two groups. Next, only the threshold (*b*) parameter estimates for the item under study were allowed to vary between the two groups. Finally, both the threshold (*b*) and slope (*a*) parameters for the item under study were allowed to vary while the remaining items continued to function as anchors.

To test for threshold-related *dif*, the fit of the model which allowed the threshold of the item to vary between groups was compared to the fully constrained model. A significant difference in fit between these two models indicated the presence of threshold-related *dif*. If this test was significant, slope-related *dif* was evaluated by comparing the fit of the model with both the slope and threshold parameters free to the model with only the threshold parameters free. If the fit of these models differ significantly, then the item exhibited slope-related *dif*. In cases where the test for threshold-related *dif* was not significant, the presence of slope-related *dif* was tested by comparing the model with the item's slope and threshold free to the fully constrained model. For the following analyses, evidence of *dif* may suggest that it is not appropriate to combine the responses of male and female beneficiaries.

Table 7 displays the tests for slope- and threshold-related *dif* and the final parameter estimates for males and females separately. When there is no evidence of *dif*, the final parameter estimates are those from the fully constrained model. If the item demonstrated threshold-related *dif*, the final parameter estimates are taken from the model which allows the thresholds to vary between the two groups.

Reviewing the results shown in Table 7 reveals that only one item, Item 46 (Medicare is offering more information), demonstrated any evidence of *dif*. The fit of the model with the thresholds free differed significantly from the fully constrained model (χ^2 (1) = 4.6, p = .03), indicating that Item 46 exhibited threshold-related *dif*. As shown in Table 7, the threshold parameter for this item has a value of 0.06 for male respondents

and a value of –0.09 for female respondents, suggesting that this item tends to be slightly more difficult for male respondents than for female respondents.

Again, no threshold-related *dif* was found for Item 49 (Can drop HMO and still be covered) when compared to the fully constrained model (χ^2 (1) = 0.5, $p > .05$). Slope-related *dif* was not found either when the model with slope and threshold parameters free was compared to the fully constrained model (χ^2 (2) = 2.5, $p > .05$).

Finally, Item 50 (HMOs cover more services) also seemed to function similarly for the two gender groups. There did not appear to be any threshold-related *dif* (χ^2 (1) = 2.1, $p > .05$) or slope-related *dif* for this item (χ^2 (2) = 2.1, $p > .05$).

Overall, the comparisons between the two genders indicated that only one item, Item 46 (Medicare is offering more information), demonstrated any form of differential item functioning for the two groups. This item showed some evidence of threshold-related *dif*, suggesting that the item was more difficult for male respondents than female respondents.

Table 7. Results of Dif Analyses Comparing Genders.

Item	Test for Threshold-Related Dif	Test for Slope-Related Dif	Males		Females	
			Slope (a)	Threshold (b)	Slope (a)	Threshold (b)
43. Plan options available	χ^2 (1) = 1.5	χ^2 (2) = 1.7	1.14	0.31	1.14	0.31
44. Medicare alone pays all expenses	χ^2 (1) = 0.5	χ^2 (2) = 1.4	1.32	-1.22	1.32	-1.22
46. Medicare is offering more information	χ^2 (1) = 4.6*	χ^2 (1) = 0.0	0.99	0.06	0.99	-0.09
47. Can report complaints about HMOs	χ^2 (1) = 0.1	χ^2 (2) = 1.3	1.23	-0.20	1.23	-0.20
48. Limited choice of doctors in HMOs	χ^2 (1) = 3.1	χ^2 (2) = 3.2	1.89	-0.26	1.89	-0.26
49. Can drop HMO and still be covered	χ^2 (1) = 0.5	χ^2 (2) = 2.5	2.60	0.21	2.60	0.21
50. HMOs cover more services	χ^2 (1) = 2.1	χ^2 (2) = 2.1	2.16	0.52	2.16	0.52

* $p < .05$

SOURCE: Centers for Medicare and Medicaid Services, Medicare Current Beneficiary Survey 1999 Supplemental BK Round-26 file.

6.2 Medicare Eligibility Status

The next set of comparisons examines possible differential item functioning in two samples, beneficiaries who are eligible for Medicare because of their age and those who are eligible for Medicare because they are disabled or have end stage renal disease. These two groups may differ greatly in their experiences with the health care system which, in turn, may have an impact on how they interpret and respond to the knowledge items. For example, beneficiaries who are disabled or have end-stage renal disease may require more medical care and therefore interact more with their health plans and the Medicare program than aged beneficiaries.

Table 8 displays the results of the *dif* analyses comparing aged and disabled/ESRD beneficiaries. Overall, the comparisons between the two eligibility groups indicated that only two items, Item 44 (Medicare alone pays all expenses) and Item 50 (HMOs cover more services) showed evidence of threshold-related *dif*.

The model fit for Item 44 (Medicare alone pays all expenses) was significantly improved by allowing the thresholds to vary between the two groups (χ^2 (1) = 15.6, p = .00008). The threshold for the aged group was –1.15 while the threshold for the disabled group was –0.80. These values suggest that Item 44 was easier for the aged group than the disabled group.

For Item 50 (HMOs cover more services), the model fit was significantly improved by permitting the threshold estimates to differ between the two eligibility groups (χ^2 (1) = 8.5, p = .004). For aged beneficiaries, the value of the threshold parameter was 0.72 and for disabled beneficiaries, the threshold parameter was equal to 0.55. These values suggest that Item 50 was more difficult for aged beneficiaries than for disabled beneficiaries.

Table 8. Results of Dif Analyses Comparing Medicare Eligibility Status Groups.

Item	Test for Threshold-Related Dif	Test for Slope-Related Dif	Aged		Disabled/ESRD	
			Slope (a)	Threshold (b)	Slope (a)	Threshold (b)
43. Plan options available	χ^2 (1) = 0.1	χ^2 (2) = 0.2	1.14	0.48	1.14	0.48
44. Medicare alone pays all expenses	χ^2 (1) = 15.6***	χ^2 (1) = 0.4	1.30	-1.15	1.30	-0.80
46. Medicare is offering more information	χ^2 (1) = 0.2	χ^2 (2) = 0.7	0.99	0.14	0.99	0.14
47. Can report complaints about HMOs	χ^2 (1) = 3.6	χ^2 (2) = 3.7	1.23	-0.03	1.23	-0.03
48. Limited choice of doctors in HMOs	χ^2 (1) = 0.9	χ^2 (2) = 0.9	1.92	-0.09	1.92	-0.09
49. Can drop HMO and still be covered	χ^2 (1) = 0.3	χ^2 (2) = 1.7	2.76	0.38	2.76	0.38
50. HMOs cover more services	χ^2 (1) = 8.5**	χ^2 (1) = 0.0	2.19	0.72	2.19	0.55

** $p < .01$, *** $p < .001$

SOURCE: Centers for Medicare and Medicaid Services, Medicare Current Beneficiary Survey 1999 Supplemental BK Round-26 file.

6.3 Managed Care Enrollment

The final set of *dif* analyses compares beneficiaries who were enrolled in a managed care plan during the past year with those who were not enrolled in a managed care plan during the past year. This comparison is particularly relevant because most of the items in the BK 7-item knowledge index address topics related to managed care plans. Beneficiaries who have been enrolled in a managed care plan could potentially have more knowledge of these plans.

Table 9 presents the results of the dif analyses comparing these two groups. The results indicated that five of the seven knowledge items (Items 44, 46, 47, 49, and 50) demonstrated threshold-related dif according to managed care enrollment. Again, no items appeared to have slope-related dif.

The model fit for Item 44 (Medicare alone pays all expenses) was significantly improved over the fully constrained model by allowing the threshold estimates to vary between the two groups (χ^2 (1) = 68.8, $p < .00001$). The threshold estimates were –2.03 for the no managed care enrollment group and –1.26 for the group with some managed care enrollment, indicating that this item was easier for those who were not enrolled in managed care during the past year.

As for Item 44, model fit was improved over the fully constrained model by allowing the thresholds for Item 46 (Medicare is offering more information) to vary between the two groups (χ^2 (1) = 61.7, $p < .00001$). The estimates for the threshold parameter were –0.86 for the no managed care enrollment group and –0.13 for the group with some managed care enrollment, suggesting that this item was less difficult for those who had not been enrolled in managed care during the past year.

As with the previous two items, Item 47 (Can report complaints about HMOs) exhibited threshold-related *dif*. The model fit for this item was improved over the fully constrained model by allowing the threshold estimates to vary between the two groups (χ^2 (1) = 19.5, $p = .00001$). The threshold estimates were –0.97 for those with no managed care enrollment and –0.60 for those with some managed care enrollment, indicating that this item was easier for those who had not been enrolled in a managed care plan during the past year.

For Item 49 (Can drop HMO and still be covered), allowing the threshold parameters to vary between the two groups significantly improved the model fit over that of the fully constrained model (χ^2 (1) = 17.5, $p = .00003$). In contrast to Items 44, 46, and 47, this item was actually more difficult for respondents with no managed care enrollment (b = -0.37) than for those with some managed care enrollment during the past year (b = -0.62).

The model fit for Item 50 (HMOs cover more services) was significantly improved by allowing the threshold estimates to vary between the two groups (χ^2 (1) = 63.7, $p < .00001$). Examining the threshold parameters for the two groups suggests that

Item 50 was more difficult for the beneficiaries who had not been enrolled in managed care (b = 0.06) than for those who were enrolled in managed care during the past year (b = -0.45).

These results indicate that five of the seven knowledge items demonstrated threshold-related *dif* according to managed care enrollment. Again, no items appeared to have slope-related *dif*. Items 44 (Medicare alone pays all expenses) and 46 (Medicare is offering more information) were less difficult for beneficiaries with no managed care enrollment during the past year. Both of these items concern the Original Medicare plan and therefore might be easier for those who have experience with the Original Medicare plan directly rather than through a Medicare managed care plan.

As might be expected, Items 49 (Can drop HMO and still be covered) and 50 (HMOs cover more services) that address topics related to managed care plans were easier for beneficiaries who had some managed care enrollment during the past year. Interestingly, Item 47 (Can report complaints about HMOs) which also addresses a topic related to managed care plans was actually more difficult for beneficiaries with some managed care enrollment. A possible explanation is that some of the beneficiaries who were not enrolled in a managed care plan during the past year may have actually left a managed care plan previously because they were dissatisfied. Dissatisfied beneficiaries might be more knowledgeable about the complaint process than those who have remained in the managed care plan.

Table 9. Results of Dif Analyses Comparing Managed Care Enrollment Groups.

Item	Test for Threshold-Related Dif	Test for Slope-Related Dif	No Enrollment		Some Enrollment	
			Slope (a)	Threshold (b)	Slope (a)	Threshold (b)
43. Plan options available	χ^2 (1) = 0.2	χ^2 (2) = 1.5	1.08	-0.35	1.08	-0.35
44. Medicare alone pays all expenses	χ^2 (1) = 68.8***	χ^2 (1) = 2.7	1.40	-2.03	1.40	-1.26
46. Medicare is offering more information	χ^2 (1) = 61.7***	χ^2 (1) = 2.6	1.05	-0.86	1.05	-0.13
47. Can report complaints about HMOs	χ^2 (1) = 19.5***	χ^2 (1) = 0.8	1.24	-0.97	1.24	-0.60
48. Limited choice of doctors in HMOs	χ^2 (1) = 1.6	χ^2 (1) = 5.2	1.86	-0.94	1.86	-0.94
49. Can drop HMO and still be covered	χ^2 (1) = 17.5***	χ^2 (1) = 2.4	2.51	-0.37	2.51	-0.62
50. HMOs cover more services	χ^2 (1) = 63.7***	χ^2 (1) = 0.9	1.95	0.06	1.95	-0.45

*** $p < .001$

SOURCE: Centers for Medicare and Medicaid Services, Medicare Current Beneficiary Survey 1999 Supplemental BK Round-26 file.

6.4 Summary of Dif Analyses

Overall, the *dif* analyses indicated that none of the items appeared to have the most detrimental type of *dif*, slope-related *dif*, suggesting that the items have the same relationship to the underlying construct for the groups. However, some of the knowledge items presented evidence of threshold-related *dif*, indicating that they were differentially difficult for certain groups.

The biggest threshold differences were found based on managed care enrollment. This result seems reasonable given that most of the items on the quiz address knowledge of managed care plans. Beneficiaries who have been enrolled in managed care plans would be expected to have more experience and presumably more knowledge of these plans than those who have not recently or have never been enrolled in them.

The implications of these results depend on the intended uses of the knowledge index. For example, in educational testing, test results can have a major impact on a student's educational future and therefore it is important to ensure that the items are not biased against a particular group. In that case, the presence of any type of *dif* suggests that the item should be removed.

However, if the intended purpose of the MCBS knowledge index is not to make a judgment about the knowledge of a particular person, but rather to measure general knowledge concerning managed care plans, then these results would suggest that perhaps these items could be retained in the quiz. In fact, discovering that some items are more difficult for certain groups may in and of itself be informative. Perhaps these groups could be targeted to receive more information about managed care options.

Decisions regarding whether to modify or discard an item should also consider the potential cause of the *dif* between the two groups. For example, in this study, it seems reasonable that differences in experience with managed care plans may account for the different difficulty levels between those who have had some enrollment and those with no enrollment in managed care plans during the past year. In fact, this hypothesis could be tested by re-analyzing the items after implementation of an education program designed to improve knowledge of managed care plans. If *dif* is no longer present, then it may suggest these items should be retained.

However, the implications of the threshold-related *dif* are less clear for the items that were found to differ according to gender or Medicare eligibility status. Perhaps these results are due to differences in comprehension of the questions or interpretation of the question wording. If this is the case, these items should be reported separately for the different groups or modified to try to eliminate the presence of *dif*. Further analyses should be conducted using other years of the MCBS in which these items were administered to determine whether this finding holds.

Finally, when making conclusions using the BK quiz items, it is important to consider that these items do appear to be easier for particular groups. In particular, caution should be used when comparing the knowledge of individuals who have been enrolled in managed care with those who have not been enrolled in managed care during the past year. Those in managed care could appear to have more knowledge than those who have not been enrolled in managed care plans recently, simply due to differences in their experiences.

7.0 Conclusions and Recommendations

Overall, the results suggested that the knowledge items included in the 1999 MCBS are good candidates for inclusion in the item pool. The factor analysis results indicated that the seven items comprised a unidimensional scale, primarily measuring knowledge of managed care plans. Because the scale contained only one factor, the items could be analyzed using conventional IRT models.

The IRT item parameters indicated that all seven items showed good discrimination ability and therefore were related to the underlying construct. Of the seven BK items, Item 46 (Medicare is offering more information) had the lowest slope while Item 49 (Can drop HMO and still be covered) had the highest slope parameter. The knowledge items had a variety of difficulty parameters, ranging from –1.27 to 0.46, suggesting that the quiz cannot effectively discriminate knowledge for beneficiaries with higher knowledge levels ($\theta > 0.46$). Adding some higher difficulty items would improve the ability of the quiz to effectively discriminate knowledge for a wider range of beneficiaries.

Examining the test information curve indicated that the 7-item quiz is most informative for measuring beneficiaries with average levels of knowledge (i.e., beneficiaries with knowledge (θ) scores ranging from –0.5 to 0.5). Based on the test standard error of measurement curve, the quiz also provides the most precise measures of knowledge for scores in this range.

In addition to evaluating the item parameters, we computed IRT-based scores for each respondent. Because both the item parameters and the individual responses are included in the calculation of the score, the IRT-based scores are more precise estimates of knowledge than scores based on Classical Test Theory.

Differential item functioning analyses were also conducted to compare the functioning of the items for three groups of beneficiaries: (1) male vs. female, (2) aged vs. disabled beneficiaries, and (3) beneficiaries with some managed care enrollment in the past year and those with none. Five of the seven items were differentially difficult according to whether respondents were enrolled in managed care during the past year.

The three quiz items contained in the 1999 BN supplement could not be analyzed using IRT because IRT can produce unstable estimates for only three items. In the future, IRT analyses could be conducted on the three BN items if either more quiz items are added to the BN supplement or if the BN items are included in the same round as the seven BK items.

A limitation of the current analyses is that during the 1999 MCBS the quiz items were administered only to beneficiaries who were in their first year of participation in the survey. Therefore, the IRT parameters may only be applicable to those who have just begun the study. There may be differences between respondents in their first year and those who have participated in the study longer. For example, responding to the survey

questions may prompt some respondents to learn more about the Medicare program and therefore the items may be easier for them than for respondents who have just started participating in the study. To determine whether the IRT parameters estimated using the 1999 data are applicable to all beneficiaries, similar analyses could be conducted with data from the BK supplement included in the 2000 MCBS which administered the seven quiz items to all MCBS participants.

The analyses in this report have identified and evaluated a set of potentially useful items for inclusion in the item pool. However, creating a large pool of items would require the development and testing of additional items. For example, while the 7-item quiz can assess knowledge of some aspects of managed care, Medicare beneficiaries may need other types of knowledge to effectively navigate the Medicare program. Therefore, additional items are required to measure understanding of the entire Medicare program.

In addition to evaluating the existing MCBS knowledge items, another goal of the current project was to develop a set of new knowledge items. Several new knowledge items have already been created and cognitively tested, addressing topics such as eligibility for and structure of Original Medicare, beneficiary rights and protections, and how to get more information and assistance.

To capitalize on the potential benefits of IRT for the MCBS, we recommend that the next step in the development of the MCBS knowledge item pool be to conduct a pre-test in which all of the newly developed items are administered to a large sample of respondents. The respondents selected for the pre-test should be representative of the population that will eventually be administered the items in the MCBS. Once the data have been collected, the IRT parameters for the items could be estimated and used to develop a set of equivalent forms that would allow different sets of respondents to receive different knowledge questions, while still receiving comparable scores. Calibration of the items would also make it possible to change the items from year to year and potentially to intersperse new items during future years.

References

Bann, C., Lissy, K.S., Keller, S., Garfinkel, S., & Bonito, A.J. (2000). *Analysis of Medicare beneficiary baseline data from the Medicare Current Beneficiary Survey: Knowledge Index Technical Note.* Report prepared for the Health Care Financing Administration. NTIS No. PB2001-102026.

Embretson, S.E., & Reise, S.P. (2000). *Item Response Theory for Psychologists.* Mahwah, NJ: Lawrence Erlbaum Associates.

Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Publications.

Hibbard, J., Jewett, J., Englemann, S., & Tusler, M. (1998). Can Medicare beneficiaries make informed choices? *Health Affairs*, *17(6)*, 181-193.

McCormack, L., Anderson, W., Daugherty, S., Ross, K., Kuo, M., & Garfinkel, S. (2000). *National Evaluation of the Medicare & You 2000 Handbook.* Report prepared for the Health Care Financing Administration. NTIS No. PB2002-100414.

Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric Theory*. McGraw-Hill: New York.

Thissen, D. (1991). *MULTILOG User's Guide: Version 6.0.* Scientific Software International, Inc.: Chicago, Illinois.

Uhrig, J.D., Squire, C., McCormack, L.A., Bann, C., Hall, P.K., An, C., & Bonito, A.J. (2001). *Questionnaire Development and Cognitive Testing Using Item Response Theory: Questionnaire Development Final Report.* Report prepared for the Centers for Medicare and Medicaid Services.

Appendix A:

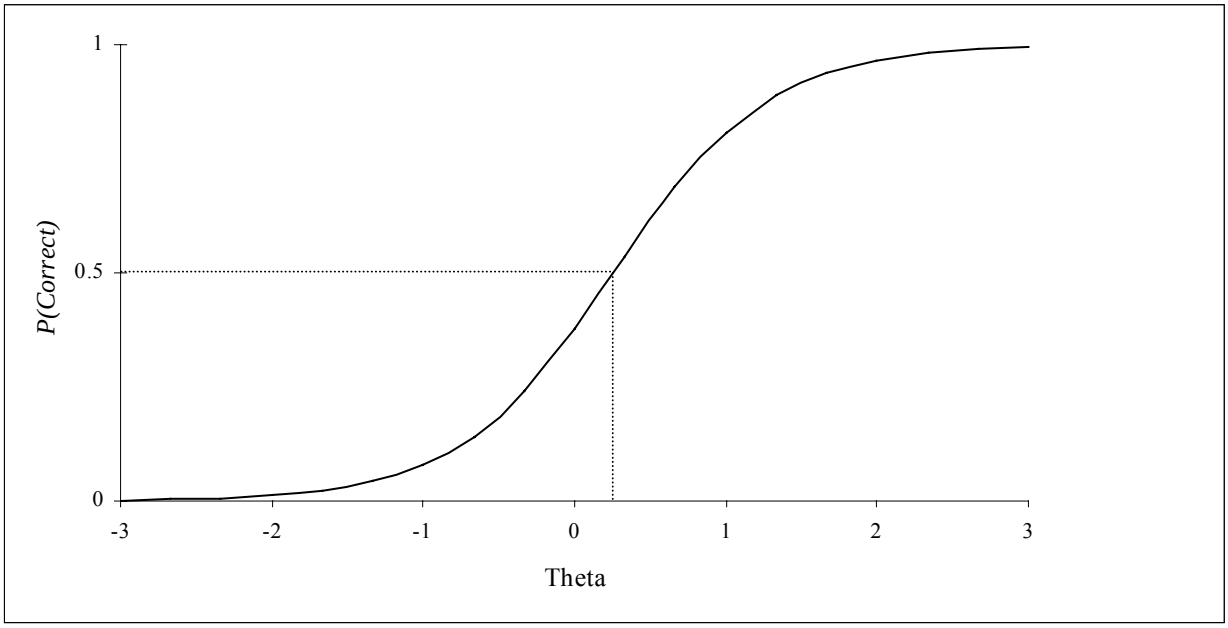
IRT Curves

Appendix A: IRT Curves

Appendix A contains graphical representations of the item response theory (IRT) analysis results described in Chapter 5. Figures A-1 to A-7 display item characteristic curves for each of the seven demonstrated knowledge items included in the 1999 MCBS Beneficiary Knowledge (BK) supplement (Round 26) and Figure A-8 presents the item characteristic curves for all seven items together. The X-axis of an item characteristic curve is labeled as theta and represents the levels of the construct. In this case, theta represents knowledge of the Medicare program. The Y-axis represents the probability of answering the item correctly. As shown in Figures A-1 to A-8, as theta increases, the probability of answering the item correctly also increases.

Figures A-9 to A-11 also include the values of theta on the X-axis. However, instead of displaying the probability of a correct response on the Y-axis as in the item characteristic curve, Figure A-9 displays the amount of psychometric information provided by each item. Figure A-10 presents the information for the entire test at each level of theta while Figure A-11 displays the test standard error of measurement.

Figure A-1. Item Characteristic Curve for Item BK 43: Plan options available (*a* = 1.14, *b* = 0.26).



SOURCE: Centers for Medicare and Medicaid Services, Medicare Current Beneficiary Survey 1999 Supplemental BK Round-26 file.

Figure A-2. Item Characteristic Curve for Item BK 44: Medicare alone pays all expenses (*a* = 1.33, *b* = -1.27).

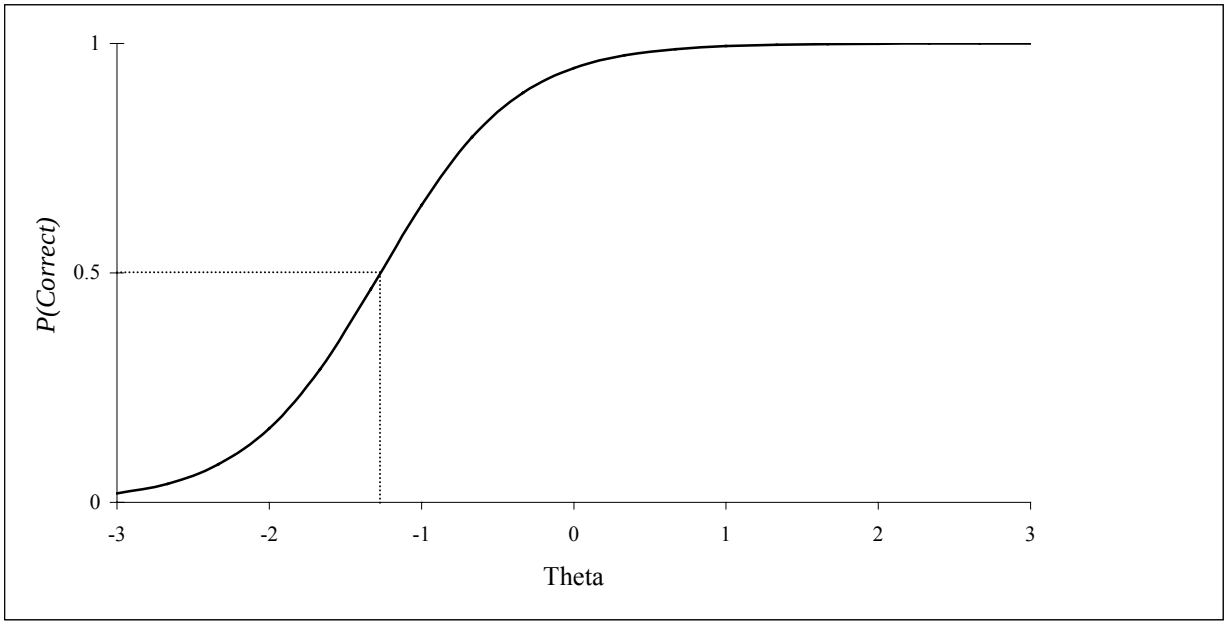


Figure A-3. Item Characteristic Curve for Item BK 46: Medicare is offering more information (*a* = 0.99, *b* = -0.08).

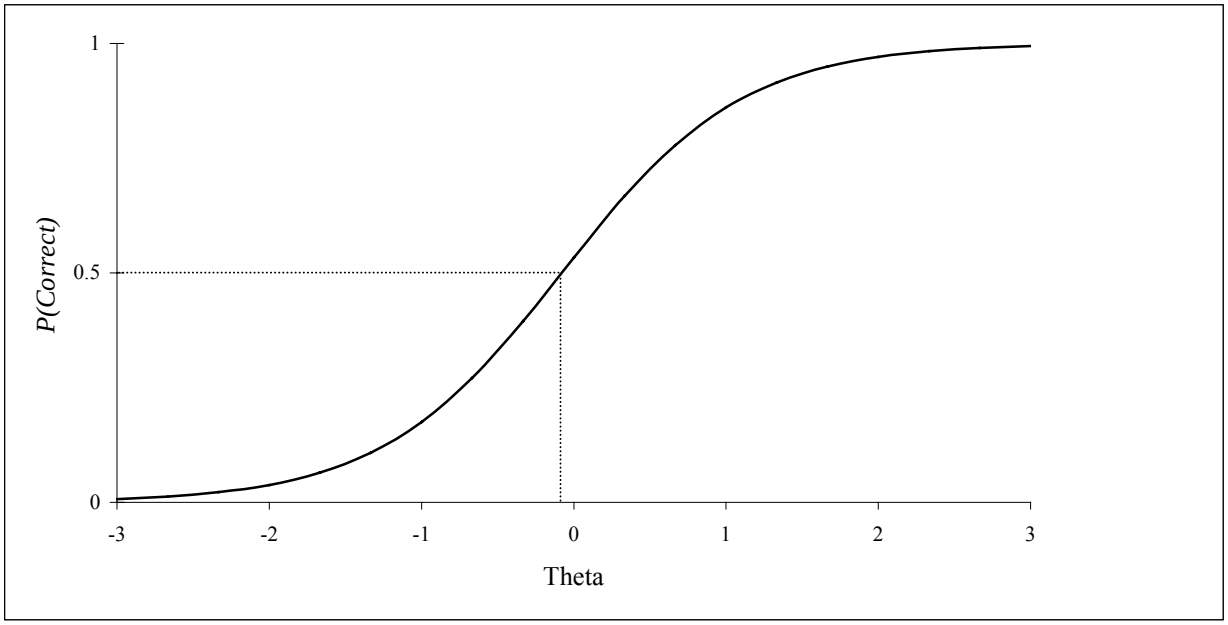


Figure A-4. Item Characteristic Curve for Item BK 47: Can report complaints about HMOs (*a* = 1.23, *b* = -0.26).

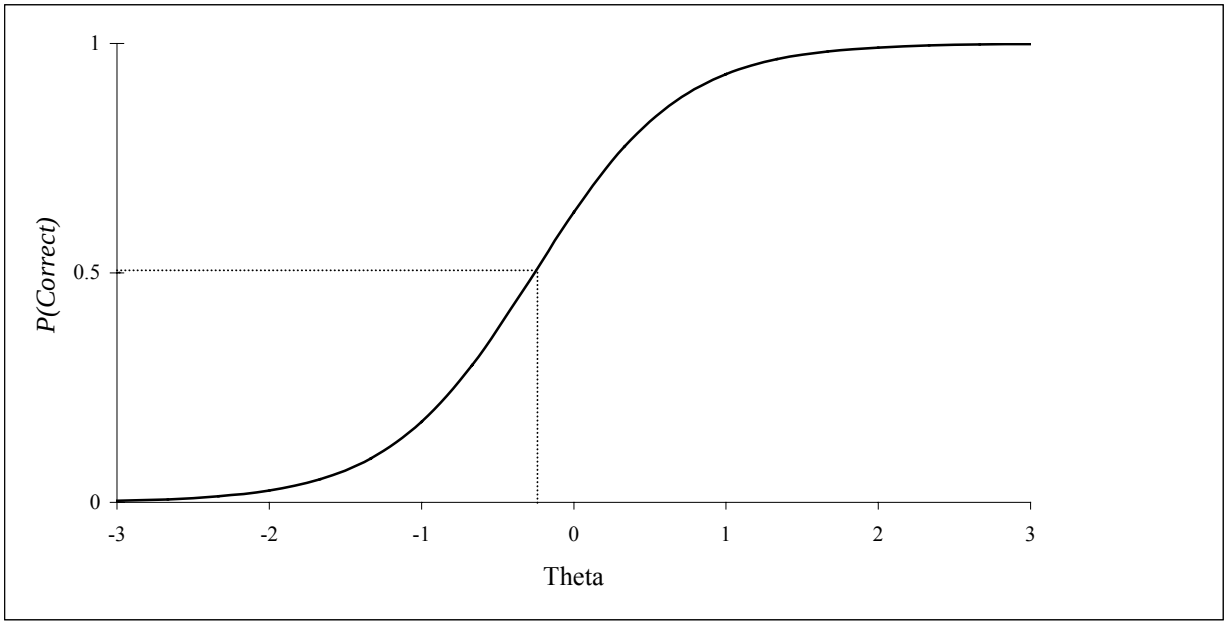


Figure A-5. Item Characteristic Curve for Item BK 48: Limited choice of doctors in HMOs (*a* = 1.89, *b* = -0.31).

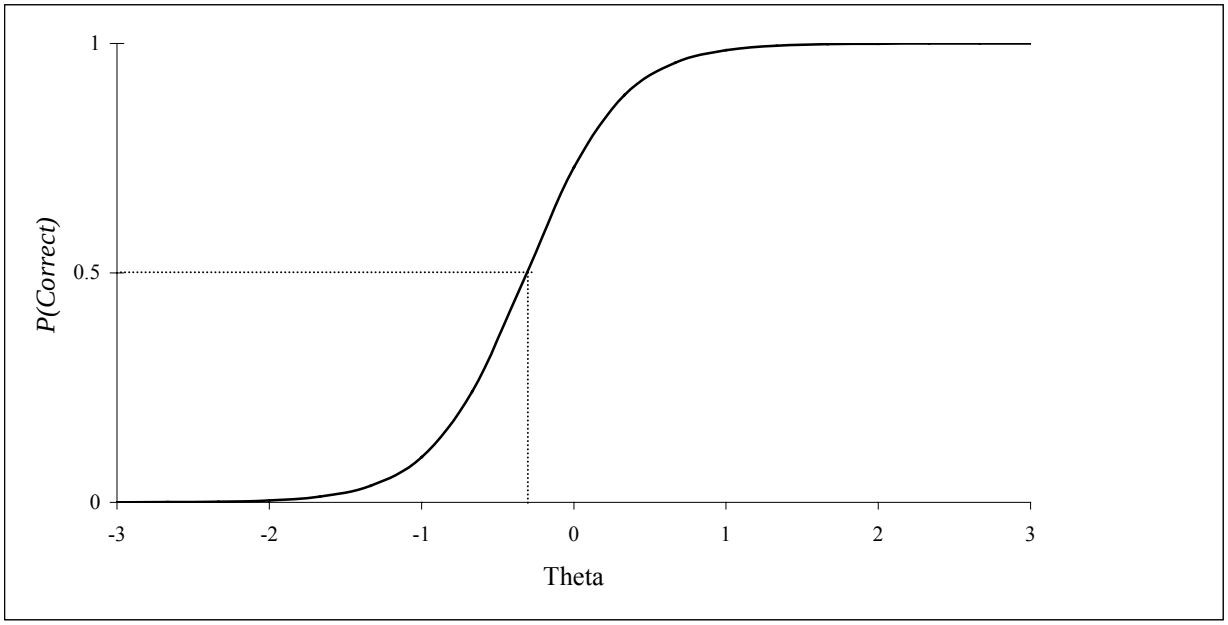


Figure A-6. Item Characteristic Curve for Item BK 49: Can drop HMO and still be covered (*a* = 2.56, *b* = 0.16).

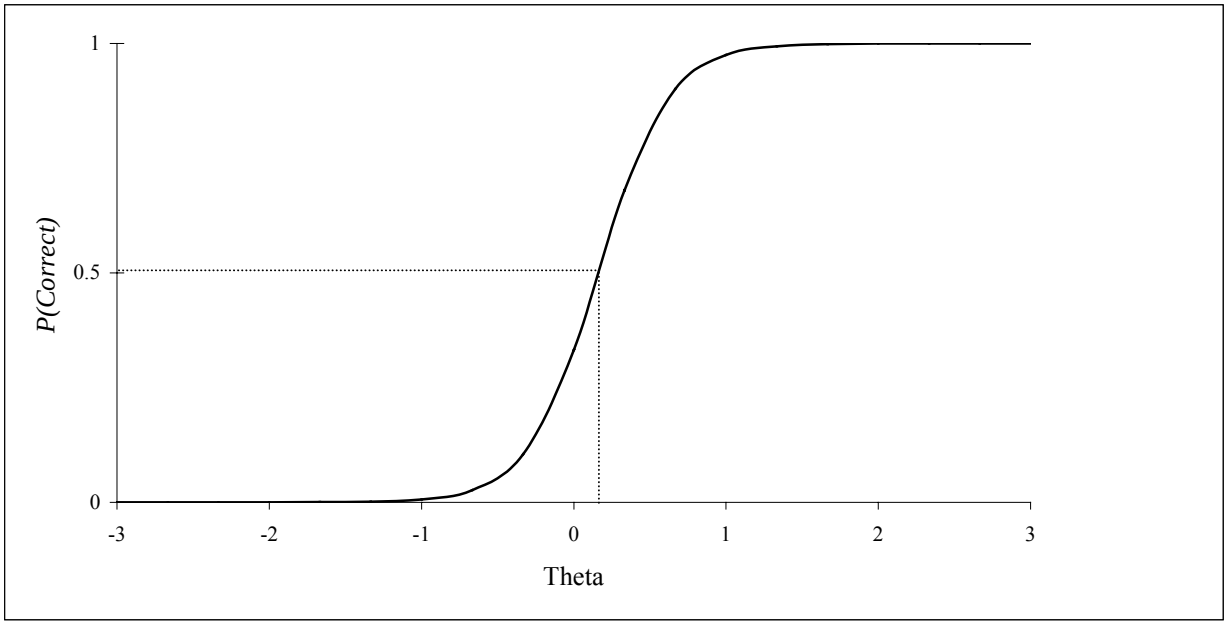


Figure A-7. Item Characteristic Curve for Item BK 50: HMOs cover more services (*a* = 2.14, *b* = 0.46).

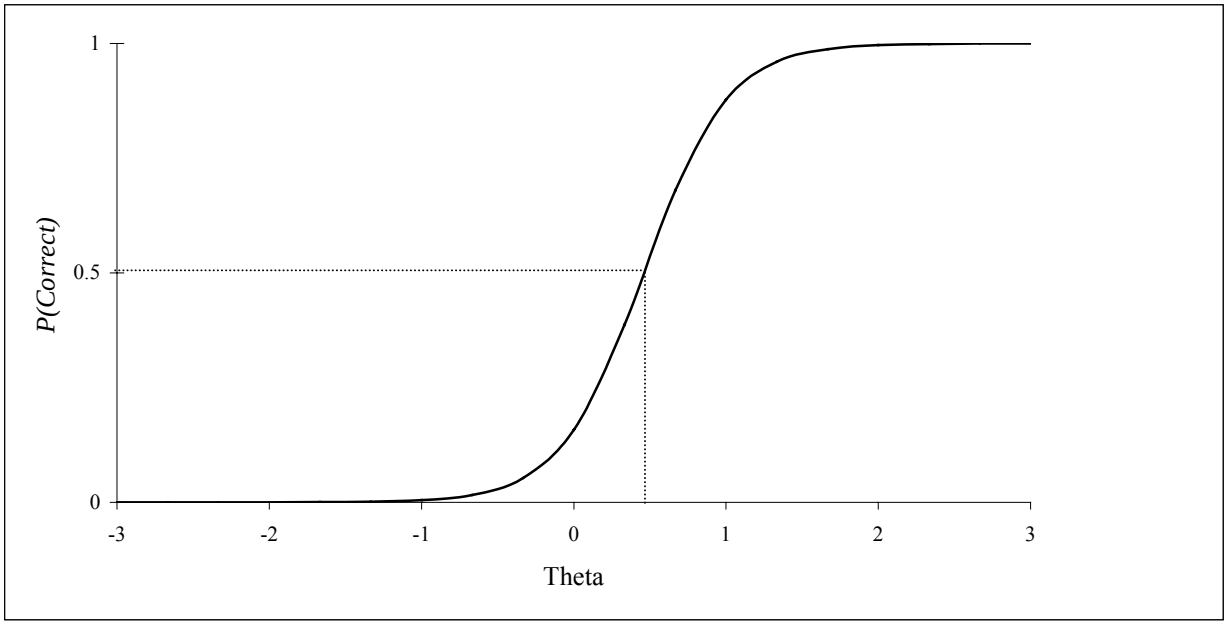


Figure A-8. Item Characteristic Curves for All Seven BK Knowledge Items.

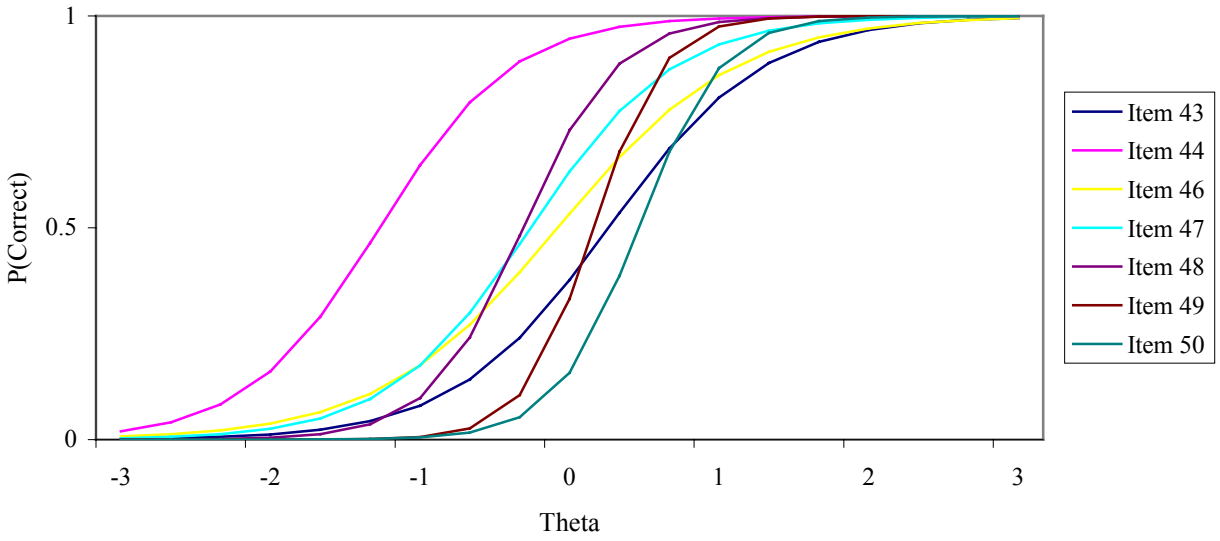


Figure A-9. Item Information Curves for All Seven BK Knowledge Items.

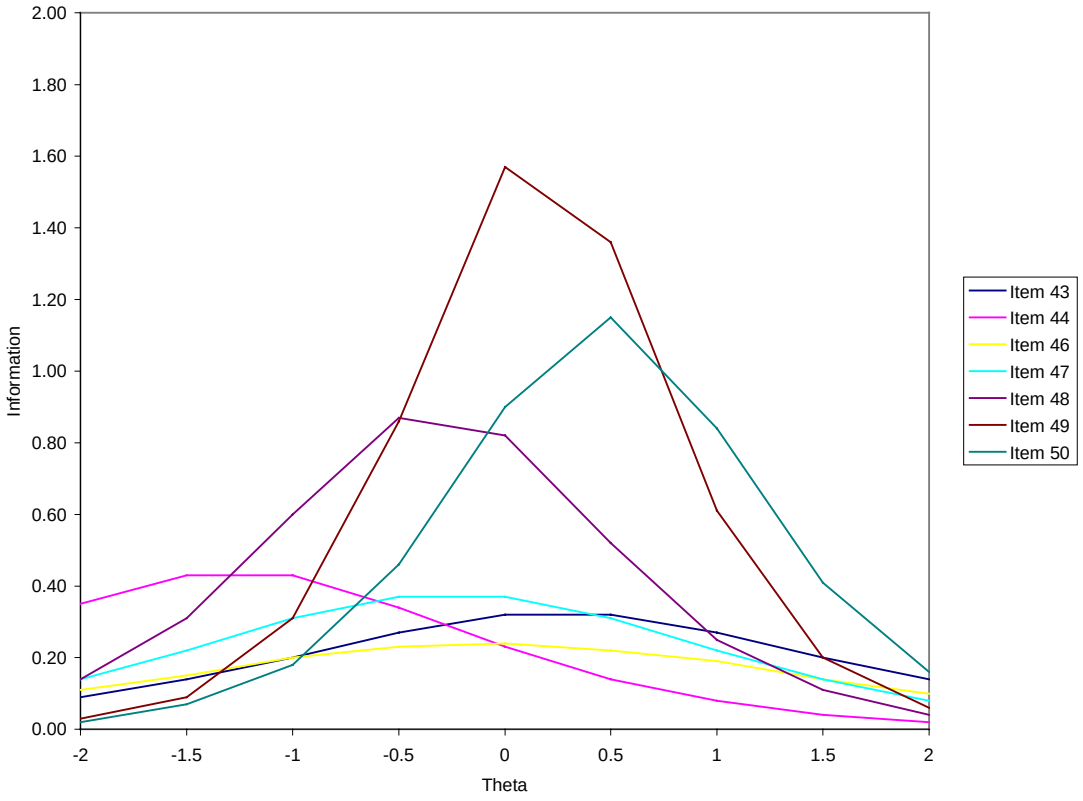


Figure A-10. Test Information Curve.

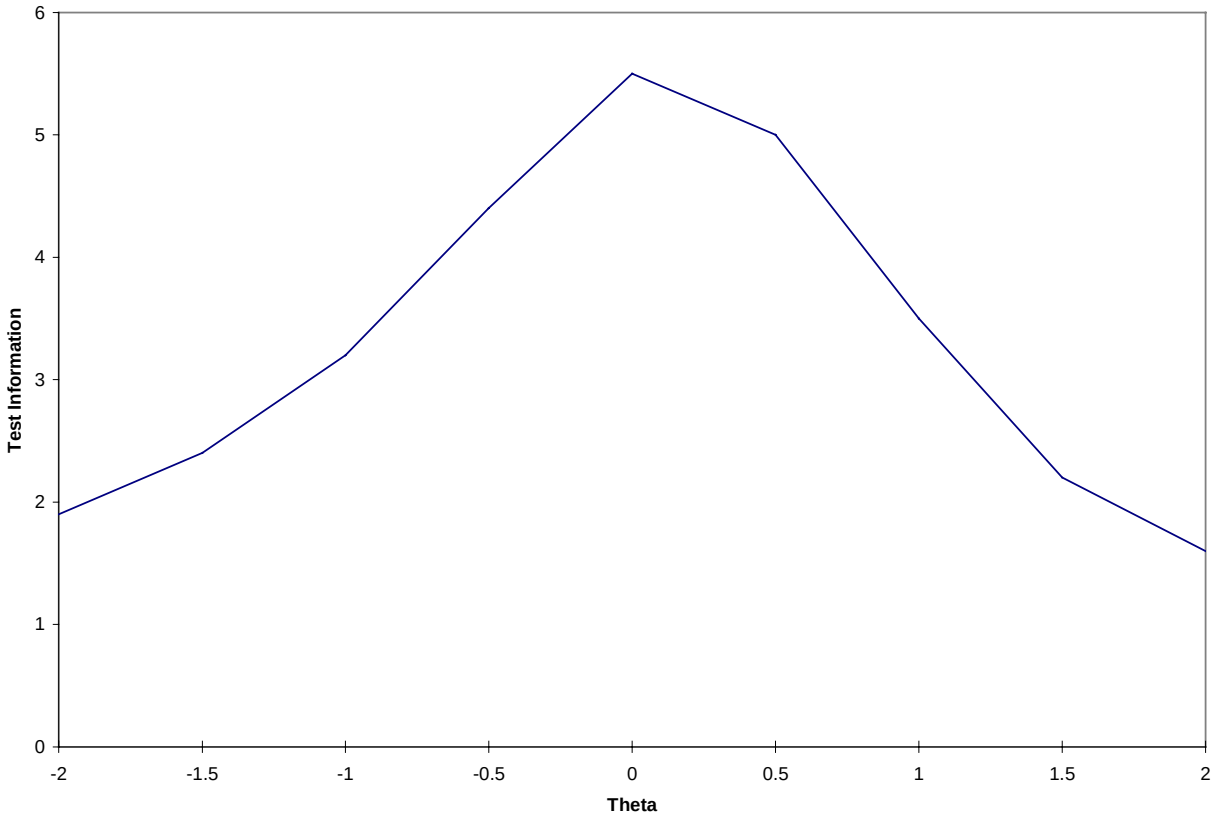
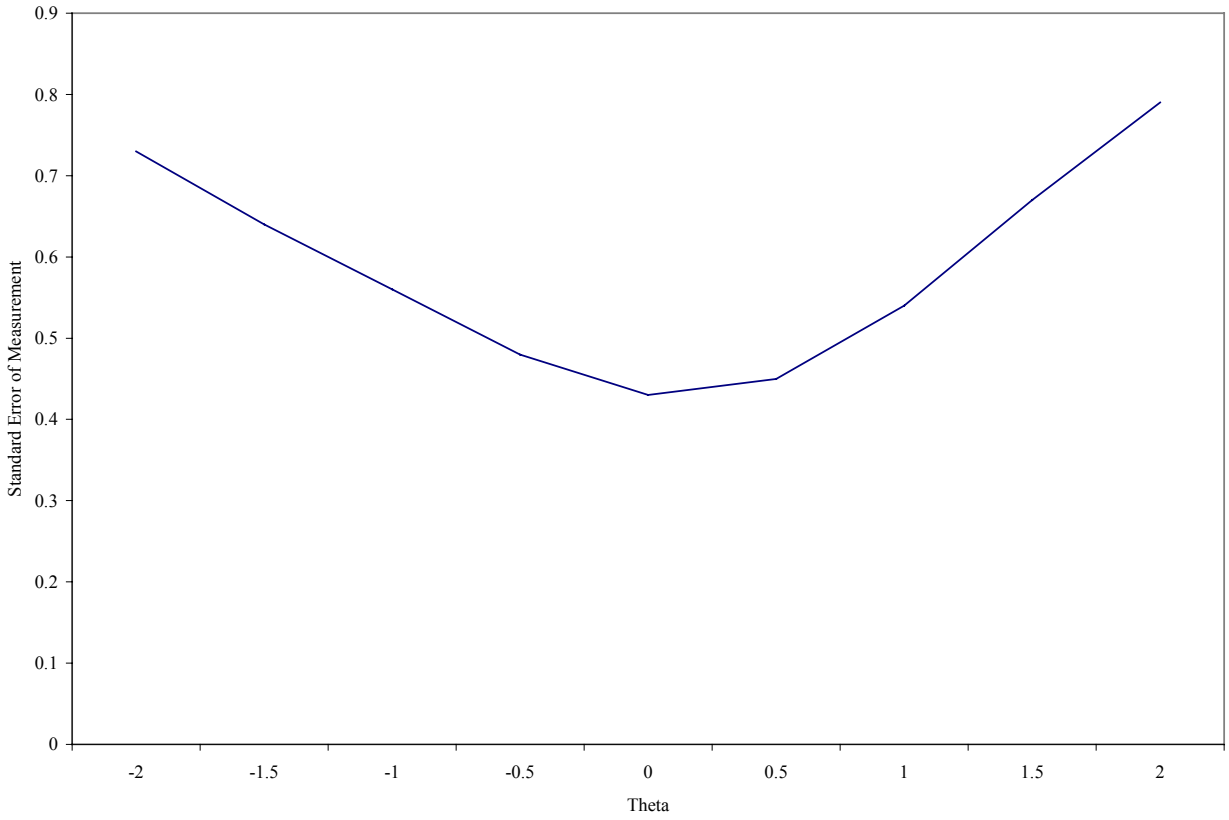


Figure A-11. Test Standard Error of Measurement Curve.



Appendix B:

IRT-Based Scores

Table B-1. List of IRT-Based Scores for All Observed Response Patterns.

Response Pattern	IRT Score	Standard Deviation
0000000	-1.56	0.64
0000001	-0.80	0.52
0000010	-0.70	0.48
0000011	**-0.26**	**0.48**
0000100	-0.88	0.54
0000101	-0.40	0.45
0000110	-0.32	0.47
0000111	0.23	0.48
0001000	-1.09	0.59
0001001	-0.53	0.44
0001010	-0.45	0.44
0001011	0.06	0.52
0001100	-0.58	0.45
0001101	-0.12	0.51
0001110	-0.01	0.52
0001111	0.47	0.41
0010000	-1.18	0.60
0010001	-0.58	0.45
0010011	0.00	0.52
0010100	**-0.63**	**0.46**
0010101	-0.18	0.50
0010110	-0.07	0.52
0010111	0.43	0.42
0011000	-0.78	0.52
0011001	-0.34	0.46
0011010	-0.24	0.49
0011011	0.30	0.46
0011100	-0.39	0.45
0011101	0.15	0.50
0011110	0.25	0.48
0011111	0.64	0.44
0100000	-1.06	0.59
0100001	-0.51	0.44
0100010	-0.43	0.44
0100011	0.09	0.51
0100100	-0.56	0.44
0100101	-0.09	0.52
0100110	0.02	0.52
0100111	0.48	0.41

Table B-1 (Continued).

Response Pattern	IRT Score	Standard Deviation
0101000	-0.70	0.48
0101001	-0.26	0.48
0101010	-0.16	0.51
0101011	0.37	0.43
0101100	-0.32	0.47
0101101	0.23	0.48
0101110	0.32	0.45
0101111	0.71	0.48
0110000	-0.76	0.51
0110001	-0.31	0.47
0110010	-0.22	0.50
0110011	0.32	0.45
0110100	-0.37	0.45
0110101	0.17	0.50
0110110	0.27	0.47
0110111	0.66	0.45
0111000	-0.50	0.43
0111001	0.00	0.52
0111010	0.11	0.51
0111011	0.54	0.41
0111100	-0.07	0.52
0111101	0.43	0.42
0111110	0.50	0.41
0111111	0.99	0.57
1000000	-1.12	0.60
1000001	-0.55	0.44
1000010	-0.47	0.44
1000011	0.04	0.52
1000100	-0.60	0.45
1000101	-0.14	0.51
1000110	-0.03	0.52
1000111	0.45	0.41
1001000	-0.74	0.50
1001001	-0.30	0.47
1001010	-0.20	0.50
1001011	0.33	0.45
1001100	-0.36	0.46
1001101	0.18	0.50
1001110	0.28	0.47
1001111	0.67	0.46

Table B-1 (Continued).

Response Pattern	IRT Score	Standard Deviation
1010000	-0.81	0.52
1010001	-0.35	0.46
1010010	-0.26	0.48
1010011	0.28	0.46
1010100	-0.41	0.44
1010110	0.23	0.48
1010111	0.62	0.44
1011000	-0.53	0.44
1011001	-0.05	0.52
1011010	0.06	0.52
1011011	0.51	0.41
1011100	-0.12	0.51
1011101	0.39	0.43
1011110	0.47	0.41
1011111	0.93	0.56
1100000	-0.72	0.49
1100001	-0.28	0.48
1100010	-0.18	0.50
1100011	0.35	0.44
1100100	-0.33	0.46
1100101	0.21	0.49
1100110	0.30	0.46
1100111	0.69	0.47
1101000	-0.47	0.44
1101001	0.04	0.52
1101010	0.15	0.50
1101011	0.57	0.42
1101100	-0.03	0.52
1101101	0.45	0.41
1101110	0.52	0.41
1101111	1.04	0.58
1110000	-0.51	0.44
1110001	-0.03	0.52
1110010	0.09	0.52
1110011	0.52	0.41
1110100	-0.09	0.52
1110101	0.41	0.42
1110110	0.48	0.41
1110111	0.96	0.56
1111000	-0.26	0.48

Table B-1 (Continued).

Response Pattern	IRT Score	Standard Deviation
1111001	0.28	0.46
1111010	0.37	0.43
1111011	0.77	0.50
1111100	0.23	0.48
1111101	0.62	0.44
1111110	0.71	0.47
1111111	1.41	0.64