

TR 2008 - 13

**Ohio Graduation Tests
Standard Setting Report
Reading and Mathematics**

Michael B. Bunch

Measurement IncorporatedTM

August 6, 2008

Revision	Description	Revised by	Approval	Date
Int. Rel.	Edited by T. Moore	TWM		8.6.08

TR 2008 – 13

Ohio Graduation Tests Standard Setting Report Reading and Mathematics

**Michael B. Bunch
Measurement Incorporated™**

June 4, 2004

Edited by Terrence Moore of the Ohio Department of Education
22 February 2006

Table of Content

Executive Summary	5
Impact by Group	6
Results by Rounds.....	8
Issues and Concerns	10
Conclusion	10
OGT Standard Setting Report	11
Overview and Training	11
Round 1	15
Round 2	20
Round 3	25
Final Results	26
Impact	30
Readiness and evaluations.....	34
Comments.....	35
Follow-up Activities	36
Conclusions	37
Appendix A: Standard Setting Plan	38
Plan Updates.....	39
Appendix B: Sample Standard Setting Materials	71
Follow up letter	72
Agenda.....	73
Outline for Round 1	74
Outline for Round 2	76
Outline for Round 3	76
Readiness Form.....	77
Workshop evaluation sheet	78
Schedule	80
Materials List	82
Ohio Graduation Test Achievement Level Descriptors	83
Appendix C: Procedures For Translating Rasch Step Values Into Achievement Levels	76
Appendix D: Report to the State Board of Education	83
Appendix DA: List of Standard Setting Committee Members	89
Appendix DB: Performance Level Descriptors	92
Appendix DC: Committee Recommended Cut Scores and Percentages Race and Gender	95

Ohio Graduation Tests Reading and Mathematics Standard Setting Report Executive Summary

The Ohio Graduation Tests (OGT) were mandated by Ohio Senate Bill 1 (SB1) of 2001, with modifications per Ohio House Bill 3 (HB3) of 2003. Ohio law calls for tests based on rigorous content and five performance levels: Advanced, Accelerated, Proficient, Basic, and Limited. Students entering ninth grade in 2003 (class of 2007) will be required to obtain a score that places them at Proficient or better in order to graduate. In the spring of 2004, Ohio tenth graders took the Reading and Mathematics tests. On April 27-29, 2004, the Ohio Department of Education convened two groups to review these tests and recommend performance standards (cut scores) for the five levels. These two groups consisted of 23 participants in Mathematics and 24 participants in Reading. In each group were classroom teachers, administrators, parents, and business representatives. Classroom teachers included both in-subject and across-subject teachers.

One of the participants in the Reading group had to leave after Round 1 due to a family emergency.

Measurement Incorporated (MI), the contractor responsible for developing the OGT, presented a standard-setting plan to the ODE and its Technical Advisory Committee (TAC) in October 2003. The ODE approved the plan, with some modifications arising from a subsequent TAC meeting in February 2004. Implementing the final, approved version of the plan, MI staff (Dr. Michael Bunch and Dr. Elliot Inman) introduced the groups to a method commonly referred to as the Bookmark Procedure (Mitzel, et al.; 2001) for setting performance standards. The two groups analyzed the tests and recommended for each test a set of cut scores to differentiate among the five groups of students using performance level descriptors (PLDs) that had been developed by the ODE with representative groups of Ohio educators and stakeholders. After three rounds of deliberation, the groups recommended cut scores.

Prior to arriving at these recommended cut scores on April 29, each group first took the Mathematics or Reading test, discussed at some length the criteria by which the open-ended items on each test were scored (six on each test), scored their own tests, reviewed and discussed PLDs, received detailed instructions in the application of the bookmark procedure, and completed a practice exercise. Prior to each round of placing bookmarks, raters completed a readiness form indicating that they fully understood the tasks they were to perform. The raters then participated in three separate rounds of identifying the place in a difficulty-ordered test booklet at which a minimally Basic, Proficient, Accelerated, or Advanced student would cease to have at least a 50-percent chance to answer the next item correctly.

At each of these points, raters placed a bookmark. Because item difficulty and student achievement level had been plotted on the same scale, the bookmark placements actually

identified not only the most difficult item a student at the threshold of a particular category could answer but the achievement level of that student as well. These individual achievement level estimates were then used to calculate a mean achievement level for each threshold, which corresponded to a unique raw score on the test. These threshold raw scores were the cut scores. They are shown, along with the associated percentages of students at each level in the spring of 2004 (class of 2006) in Table ES-1. The percentages shown in Table ES-1 are preliminary estimates based on a sample of 74,337 reading tests and 76,092 mathematics tests out of approximately 150,000 students who took the tests in March 2004.

Table ES-1
Final Cut Scores and Percentages of Students in Each Category

Category	Mathematics		Reading	
	Cut Score (Out of 46)	Percent in Category	Cut Score (Out of 48)	Percent in Category*
Limited	---	17.2	---	12.1
Basic	13.5	14.9	13.5	10.3
Proficient (Graduation)	19.0	31.3	20.0	25.9
Accelerated	30.0	19.6	31.5	25.9
Advanced	37.5	17.0	39.0	25.9
Proficient or Above		67.9		77.7

Notes: may not sum to exactly 100% due to rounding.

After the first round of bookmark placements, raters were allowed to see where the other members of their groups had placed their bookmarks and how those bookmarks translated into raw cut scores. They also examined tables and graphs showing percentages of students at each score point. This information was used to point out how many students would be classified at each level if the first round recommendations held. Raters examined not only overall score distributions but distributions by race and gender as well and discussed the implications at length.

Raters were given a second opportunity to place bookmarks dividing the five groups of students. Once again, group facilitators from MI shared with the groups the results and led discussions of the implications. Raters then entered a third set of ratings, indicating not only the pages at which the bookmarks should be set, but the corresponding raw cut scores and associated percentages of students scoring at or above that cut score. The final ratings are shown in Table ES-1.

Impact by Group

At each round, raters examined not only the numbers of students who would be classified into each level overall but by race and gender as well. These final distributions across the five levels are shown in Tables ES-2 and ES-3. Because some students failed to indicate race or gender (or either), the numbers do not add up to the totaled sample size of 74,337 for Reading or 76,092 for Mathematics. The racial groups included the following:

American Indian (AmInd), Asian or Pacific Islander (As-PI), Black or African-American (BL-AA), Hispanic (Hisp), Multiracial (Multi), Other (Other), White (White). It should also be noted that the results shown in Tables ES-1, ES-2, and ES-3 are based on the samples of students for whom complete data were available at the time of the standard setting; i.e., the data that were shared with the raters. The final data sets, which did not vary significantly from the sample data sets, are included in Appendix C of this report.

Table ES-2
Distribution of Students by Race and Sex: Mathematics
(Entries are percentages.)

Group (Tested)	Category					Proficient or Above
	Limited	Basic	Proficient	Accelerated	Advanced	
AmInd (139)	28.78	10.79	34.53	12.95	12.95	56.95
As-PI (881)	8.06	10.22	24.29	21.45	35.98	82.50
BL-AA (10,239)	36.43	24.65	28.59	7.82	2.51	40.08
Hisp (1,499)	30.82	22.08	29.62	10.81	6.67	45.16
Multi (940)	23.40	16.49	30.43	16.49	13.19	60.11
Other (204)	28.92	15.20	35.29	7.84	12.75	55.88
White (58,986)	13.20	12.93	31.95	22.19	19.72	73.86
Total by Race 72,888	16.97	14.78	31.33	19.80	17.12	68.25
Female (37,395)	14.40	15.60	34.04	20.19	15.77	70.00
Male (38,487)	19.86	14.17	28.61	19.14	18.22	65.97
Total by Gender 75,882	17.17	14.87	31.29	19.66	17.01	67.96

Notes: 1. May not sum to exactly 100% due to rounding.
2. Some examinees failed to report Race or Gender.

Table ES-3
Distribution of Students by Race and Sex: Reading
(Entries are percentages.)

Group (Tested)	Category					Proficient or Above
	Limited	Basic	Proficient	Accelerated	Advanced	
AmInd (118)	21.19	16.95	26.27	24.58	11.02	61.87
As-PI (871)	8.61	7.12	20.67	25.49	38.12	84.28
BL-AA (10,189)	23.88	17.73	33.07	17.35	7.96	58.38
Hisp (1,508)	22.08	18.30	28.12	20.29	11.21	59.62
Multi (964)	14.73	13.07	27.07	23.55	21.58	72.2
Other (197)	22.84	12.69	29.44	21.32	13.71	64.47
White (57,318)	9.38	8.64	24.43	27.74	29.81	81.98
Total by Race 71,165	11.84	10.22	25.75	25.99	26.20	77.94
Female (36,470)	7.28	8.68	26.32	27.37	30.35	84.04
Male (37,678)	16.58	11.90	25.41	24.48	21.63	71.52
Total by Sex 74,148	12.01	10.32	25.86	25.90	25.92	77.68

Notes: 1. May not sum to exactly 100% due to rounding.
2. Some examinees failed to report Race or Gender.

Results by Rounds

As noted above, raters engaged in three rounds of setting bookmarks to derive cut scores. The placements, particularly for the Proficient category, were extremely stable across rounds. Results by round are shown in Tables ES-4 and ES-5. Percentages of students in each group are shown in parentheses. As with the first three tables, the numbers shown in Tables ES-4 and ES-5 reflect the data that were available on April 27, 2004. Complete data sets are included in Appendix C.

Table ES-4
Cut Scores by Round: Mathematics
(Percentages of students in each group are shown in parentheses.)

Category	Round		
		2	3
Limited	--- (15.0)	--- (17.8)	--- (17.2)
Basic	13.0 (17.1)	14.0 (14.3)	13.5 (14.9)
Proficient (Graduation)	19.0 (24.8)	19.0 (30.1)	19.0 (31.3)
Accelerated	27.5 (14.4)	29.5 (19.2)	30.0 (19.6)
Advanced	33.0 (28.7)	37.0 (18.6)	37.5 (17.0)
Percent of Students Proficient or Above	67.9	67.9	67.9

Notes: may not sum to exactly 100% due to rounding.

Table ES-5
Cut Scores by Round: Reading
(Percentages of students in each group are shown in parentheses.)

Category	Round		
	1	2	3
Limited	--- (11.0)	--- (11.3)	--- (12.1)
Basic	13.0 (11.4)	13.0 (11.4)	13.5 (10.3)
Proficient (Graduation)	20.0 (16.5)	20.0 (24.4)	20.0 (25.9)
Accelerated	28.0 (26.0)	31.0 (23.6)	31.5 (25.9)
Advanced	36.5 (35.1)	38.0 (29.6)	39.0 (25.9)
Percent of Students Proficient or Above	77.6	77.6	77.7

Notes: may not sum to exactly 100% due to rounding.

As shown in Tables ES-4 and ES-5, a considerable amount of group consensus building was achieved between Rounds 1 and 2. From Round 2 to Round 3, there was some movement by outlying raters, but not as much as there had been from Round 1 to Round 2. It is also noteworthy, particularly for the Proficient (Graduation) category, that the group means did not vary at all across rounds and that all positions had essentially solidified by the end of Round 2. Because the cut scores for Proficient (the Graduation standard) did not change from the first round, the percentages of students in the Proficient category or above did not change. However, because the cut scores for Accelerated and Advanced shifted upward by rounds, the distributions of students in the top three categories changed, typically resulting in more students in the Proficient category and fewer in the Accelerated and Advanced categories.

Issues and Concerns

After Round 1, ODE and contractor staff met to review the results of Round 1. The primary issues were the raw cut score for Proficient, the percentages of students in the Proficient category, and the percentages of students in the Advanced category, relative to the percentages in Proficient and Accelerated. Prior to Round 2, Dr. Bunch and Dr. Inman led discussions with their respective groups, during which they made sure these issues were addressed.

Raters in both groups discussed these issues at length. Both groups concluded that the relatively low cut scores for Proficient (Graduation) appeared unusual but were reasonable representations of the content a minimally proficient student, on track for graduation, could be expected to know. The firmness of this belief was evidenced in Rounds 2 and 3. The cut scores for Proficient (Graduation) did not change by even half a raw score point after the first round. Raters did, however, temper their judgments of the Accelerated and Advanced categories, noting that large numbers of students in the highest category, relative to Proficient and Accelerated, might be unrealistic. In Rounds 2 and 3, they re-examined the definitions for Accelerated and Advanced and in many instances applied more restricted interpretations to the Advanced category. Those reinterpretations were more evident in Round 3 than in Round 2. In the end, there were fewer students in the Advanced category than in Proficient or Accelerated in Mathematics. In Reading, the percentages of students in the top three categories were essentially the same (about 26 percent in each).

Conclusion

Standard setting is a combination of art and science. It combines the democratic process of group interaction and decision making with carefully planned and executed steps based on a well-defined mathematical model. The process by which the standard-setting activity for the Reading and Mathematics tests of the OGT was carried out, described in detail in the body of this report, was meticulously crafted by experienced psychometricians and reviewed by a national body of experts in the field. The plan was carried out under the supervision of ODE staff and an external reviewer who is also an expert in this field. Because there are no “true” cut scores for any test, the recommended cut scores are only as valid as the process by which they were derived. By all accounts, this standard-setting activity was well planned and executed. The recommended cut scores should be taken as valid.

The 46 participants in this activity – the Ohio educators and stakeholders invited to review the tests and provide the ratings – voiced overwhelming support for the process and outcomes. A complete analysis of their responses to 14 specific evaluation statements, along with their comments, is included in the final section of this report.

OGT Standard Setting Report

On May April 27-29, 2004, Dr. Michael Bunch and Dr. Elliot Inman of Measurement Incorporated (MI) met with groups of Ohio educators, parents, and community representatives to conduct standard-setting activities for the Ohio Graduation Tests in Reading and Mathematics. The standard-setting activities were conducted in accordance with a plan developed by Drs. Bunch and Inman and approved by the Ohio Department of Education (ODE) and the Ohio Technical Advisory Committee (TAC). That plan is included as Appendix A to this report.

This report documents the three-day standard-setting process and ensuing cut score recommendations of the participants. It also provides a record of the fidelity of that process to the plan submitted to the TAC at its October 2003 meeting. Additional details are provided in six appendices:

- A Standard Setting Plan
- B Standard Setting Materials
- C Impact Data
- D Evaluations by Raters
- E Technical Details of the Bookmark Procedure
- F Secure Documents

Overview and Training

Day 1 (AM). On Tuesday, April 27, Dr. Bunch opened the three-day session at 8:30 AM and introduced the MI staff and an independent observer, Dr. Thomas Hirsch of Assessment and Evaluation Services (AES). He then introduced Ms. Jan Crandell of the ODE, who introduced other ODE staff and invited participants to introduce themselves. Participants included reading and mathematics teachers as well as administrators, parents, and business and community representatives. A summary of participant characteristics is included in Table 1. A list of all participants is included in Appendix B

Ms. Crandell provided an overview of the testing program, emphasizing its history and the impact of Senate Bill 1 as well as the rationale for assembling those present. Her presentation is included in Appendix B. She then invited Dr. Bunch to present the plan for the three-day period and orient participants to the tasks they would perform. Dr. Bunch presented an overview of the standard-setting process as well as the test-development process, emphasizing the involvement of teachers and community members at several points in the process. Dr. Bunch also stressed the role of the participants as advisory to the ODE and Ohio Board of Education, which will make a final decision on cut scores after considering the recommendations of these groups as well as the TAC and the Test Steering Committee (TSC). His presentation is also included in Appendix B. There were several questions during both presentations, and either MI or ODE staff answered all to the satisfaction of the questioners.

Table 1
Summary of Participant Characteristics

Category	Mathematics	Reading	Total
Female	15	14	29
Male	8	10	18
African American	3	3	6
Asian/Pacific Islander	1		1
Hispanic		1	1
White	17	20	37
Other Race	2		2
Teacher	15	16	31
Higher Education	1	1	2
School Board	1		1
District/School Administrator	1	3	4
Business/Community Representative	4	3	7
Not specified	1	1	2
Total Members	23	24	47

At 10:15 a.m., participants adjourned to their separate meeting rooms and prepared to take the spring test that students had just completed. For the rest of the day, the Reading and Mathematics groups remained separated from each other, with Dr. Inman leading the Mathematics group, and Dr. Bunch leading the Reading group. Overview and directions for test administration took approximately 15 minutes. Testing started at about 10:30 AM for both groups and ended at 12:30 PM. After a one-hour break for lunch, participants reconvened to score their tests, using scoring keys and guides prepared by MI. These keys and guides are included in a separate collection of secure documents.

Day 1 PM. The scoring guides consisted of scoring rubrics for each of the open-ended items in the tests, followed by examples of student responses at each score point. Some of the student responses were annotated, and some were not. Drs. Bunch and Inman discussed the rubrics and sample papers for each item and then directed participants to score their own responses by comparing them to the rubrics and sample responses. Both Dr. Bunch and Dr. Inman stressed the fact that the training they were providing for scoring the open-ended responses was far less detailed than that provided for MI scorers. MI scorer training lasts several days and includes multiple rounds of scoring of sets of student responses, ending with one or more qualifying rounds in which potential scorers must match scores for a range of student responses previously scored by MI senior scoring leaders. Those who fail to match the specified number of scores are retrained or released from the project. The purpose of the training that standard-setting participants

received was not to make them professional scorers but to give them just enough insight into the process to score their own responses.

After participants had scored their tests, they took a short break and reconvened to examine and discuss achievement level definitions (also called Performance Level Descriptors). These definitions, developed by the ODE, describe in general terms the typical skills and proficiencies of students at the Limited, Basic, Proficient, Accelerated, and Advanced levels of achievement. These definitions are included in Appendix B. Dr. Bunch led a discussion in the Reading group, while Dr. Inman led a discussion in the Mathematics group. Participants were encouraged to add clarifying statements to their definitions for use later in the standard-setting process.

Having studied and discussed the definitions for all five achievement levels, participants were then encouraged to think of the student who would just barely qualify as Basic, Proficient, Accelerated, and Advanced. Drs. Bunch and Inman emphasized the fact that each achievement level is a broad band with a considerable difference between students at the high end and those at the low end of the band. By focusing on the just barely Basic, Proficient, Accelerated, or Advanced student, participants would be identifying the starting point for each of these achievement levels; hence, they would be identifying the cut points. The discussions of achievement levels ended at approximately 4 PM for both groups. Drs. Bunch and Inman collected and accounted for all secure materials and adjourned for the day.

Day 2 AM. On Wednesday, April 28, participants convened at 8:30 a.m. for an orientation to the bookmark standard-setting procedure. Dr. Bunch led the orientation for the combined groups. That presentation is included in Appendix B, while technical details associated with the construction of the difficulty-ordered test booklet are included in the standard setting plan (Appendix A). During the presentation, Dr. Bunch reminded participants of the discussions the previous day, emphasizing the notion of “just barely” Basic, Proficient, Accelerated, and Advanced. Throughout the discussions, Drs. Bunch and Inman responded to questions and helped clarify the task to be performed and the focus of the bookmark procedure:

What knowledge or skill is required to earn this point?
What makes later items more difficult than earlier items?
Think of a large group of students at the cut score for this level. Would about 50% of them earn this point?

At the end of the presentation, participants received a practice test of six items (representing a total of 10 points), including both multiple-choice and open-ended items. They were instructed to begin on page 1 of the difficulty-ordered test booklet, consider the barely proficient student, and apply the three questions above to each item. At the point at which they were no longer able to answer “Yes” to the final question, they were to go back to the previous question (i.e., the last one for which they could answer “Yes”) and place a bookmark. The practice test booklets are included in a separate document

that contains all test booklets and other secure materials. A sample page is shown in Figure 1.

The large number in the upper right corner of the page is the page number in the difficulty-ordered booklet. The item in this example would have been on the first page. The information following the difficulty-ordered page number is the original item number and Rasch-based achievement data showing the student achievement level required to have a 50% chance to answer this question correctly. The passage title is shown above the question because the passages are not printed in the difficulty-ordered test booklet. This is because the various items for any given passage could be scattered throughout the test booklet. Therefore, for the Reading test, participants also received a separate booklet of passages, along with the other secure materials). At the bottom of the page is the correct answer for the item.

Figure 1. Sample page from a difficulty-ordered test booklet

1

Item 2

Achievement level required for a 50% chance to answer correctly: -1.363

Secret Gifts

Which of these best supports the idea that Mary McLeod Bethune is concerned with helping young people find their way in the world?

- A. the legacy she leaves in her will
- B. her desire to return and help Essie
- C. her zeal for her own place in history
- D. the way she inspires Essie to believe

Answer: A

Had this been an open-ended item, the question or prompt would have been followed by a sample response at a particular score point. In a full difficulty-ordered test booklet, each such item would appear once for each of its score points, twice for a 2-point item and four times for a 4-point item. Since it is more difficult to receive a 4 than a 1 (i.e., a score

of 4 represents a higher level of student achievement than does a score of 1), the page with the “4” response would appear much later in the test booklet than the page with the “1” response. In the practice booklet, consisting of only six pages, only selected pages were included. The Mathematics practice booklet contained four multiple-choice items, one 2-point item showing a “2” response, and one 4-point item showing a “4” response. The Reading practice booklet contained a single passage with four multiple-choice items and one 4-point item. For the 4-point item, a sample response for a score point of 2 was shown, along with a sample response for a score point of 4.

After completing the practice test, participants called out their bookmarked pages, and Drs. Bunch and Inman tallied them. Participants then discussed reasons for setting bookmarks at various pages in the booklet. Finally Drs. Bunch and Inman, using the Rasch-based student achievement level printed at the top of each page, calculated the mean achievement level associated with the page numbers called out by the participants and translated this achievement level into a raw score, using the Rasch model. Details of this procedure are included in Round 1.

Participants then reviewed the procedure, their bookmarks in the practice booklets, and the achievement level definitions. Drs. Inman and Bunch asked for and responded to questions and called participants’ attention to the Readiness Form (included in Appendix B) in each participant’s packet. Participants were asked to respond to the following statement:

I have completed the practice test, and I understand what I need to do to complete Round 1.

(Circle one): **Yes** **No**

Drs. Inman and Bunch checked to make sure each participant had responded positively to the Readiness Forms. All participants had circled “Yes,” so they proceeded to Round 1 after a break for lunch. After collecting and accounting for all secure materials, Drs. Inman and Bunch dismissed their groups for lunch.

Round 1

After lunch, Drs. Bunch and Inman reminded their groups of the task before them and assigned them to teams of four (with one group of three) so that the subject-matter teachers, non-subject-matter teachers, parents, administrators, and others were evenly divided among teams, giving each team diverse points of view. Participants remained in these teams for the remainder of the session. Once they had joined their teams, each participant received a rater number. Each rater then recorded this number on each new piece of material.

Drs. Bunch and Inman distributed the difficulty-ordered test booklets, reading passage booklets (Reading group only), and a bookmark. The bookmark form is included in Appendix B, and the ordered booklets are included in a separate document with other secure materials. The bookmark has places for raters to enter three bookmarks for Round

1, Round 2 and again for Round 3. Raters were directed to reintroduce themselves to the other members of their teams and to discuss briefly their views of the test contents and the nature of students just barely in the Basic, Proficient, Accelerated, and Advanced categories. They were then directed to work alone in silence to place their bookmark for Proficient. After placing the Proficient bookmark, group members discussed their placements, made any adjustments they thought appropriate, and entered the other three bookmarks.

After all members of a team had placed the remaining three bookmarks, they discussed them. Drs. Bunch and Inman noted that the purpose of the discussion was to allow all team members to hear from others on their teams before turning in their Round 1 bookmarks but not to move the team toward a consensus. After small-group discussion ranging from ten to forty minutes, each team member turned in his or her bookmark, test booklet, and other secure materials. After accounting for all materials, Drs. Bunch and Inman dismissed individual raters for the day.

After collecting all bookmarks, MI staff entered the page numbers and associated achievement levels into Microsoft Excel spreadsheets which calculated the mean and standard deviation of the achievement levels. These mean achievement levels, along with low, mean minus one standard deviation, mean plus one standard deviation, and high were translated into raw scores using Table 20.1 of the Rasch-based WINSTEPS output. Results are shown in Tables 2 and 3 for Mathematics and Reading, respectively.

Table 2
Results for Mathematics Round 1

Rater	Basic Pg	Basic Ach	Prof Pg	Prof Ach	Acc Pg	Acc Ach	Adv Pg	Adv Ach
1	4	-1.12094	16	-0.29154	29	0.46628	38	1.02868
2	8	-0.6275	26	0.37107	39	1.04881	41	1.11953
3	9	-0.58382	22	0.127362	33	0.66094	40	1.05822
4	10	-0.51333	21	0.05466	34	0.8333	39	1.04881
5	8	-0.6275	17	-0.24375	33	0.66094	38	1.02868
6	11	-0.49392	17	-0.24375	34	0.8333	46	1.91588
7	11	-0.49392	26	0.37107	33	0.66094	40	1.05822
8	8	-0.6275	17	-0.24375	30	0.56394	39	1.04881
9	3	-1.27536	16	-0.29154	30	0.56394	38	1.02868
10	11	-0.49392	26	0.37107	36	0.93394	43	1.43293
11	6	-1.08532	11	-0.49392	30	0.56394	38	1.02868
12	10	-0.51333	18	-0.14562	35	0.83748	44	1.47756
13	4	-1.12094	13	-0.43292	27	0.44086	38	1.02868
14	8	-0.6275	26	0.37107	39	1.04881	43	1.43293
15	9	-0.58382	22	0.127362	34	0.8333	41	1.11953
16	10	-0.51333	22	0.127362	33	0.66094	38	1.02868
17	8	-0.6275	17	-0.24375	37	0.99888	40	1.05822
18	11	-0.49392	19	-0.1201	34	0.8333	42	1.18172
19	9	-0.58382	17	-0.24375	28	0.45274	39	1.04881
20	5	-1.10239	17	-0.24375	26	0.37107	37	0.99888
21	5	-1.10239	16	-0.29154	29	0.46628	38	1.02868
22	7	-0.71657	21	0.05466	35	0.83748	45	1.63433
23	2	-1.53558	12	-0.45376	28	0.45274	39	1.04881
Rasch Parameters								
Mean		-0.759		-0.087		0.697		1.169
SD		0.310		0.279		0.209		0.241
M-1SD		-1.069		-0.366		0.487		0.928
M+1SD		-0.449		0.192		0.906		1.410
Raw Score Parameters								
Mean Cut		13.00		19.0		27.5		33.0
Low Cut		8.0		15.0		24.0		31.0
Cut-1SD		10.5		16.5		25.0		30.0
Cut+1SD		15.5		22.0		30.0		35.0
High Cut		15.0		24.0		31.5		39.0

Table 3
Results for Reading Round 1

Rater	Basic Pg	Basic Ach	Prof Pg	Prof Ach	Acc Pg	Acc Ach	Adv Pg	Adv Ach
1	9	-0.91553	19	-0.26532	30	0.66312	45	1.93022
2	9	-0.91553	22	-0.12479	30	0.66312	42	1.58744
3	9	-0.91553	30	0.66312	38	1.30327	45	1.93022
4	6	-1.07021	12	-0.65801	25	-0.02999	41	1.42965
5	10	-0.90007	19	-0.26532	29	0.58512	38	1.30327
6	6	-1.07021	24	-0.06231	37	1.02657	42	1.58744
7	9	-0.91553	30	0.66312	38	1.30327	44	1.81616
8	10	-0.90007	19	-0.26532	29	0.58512	38	1.30327
9	9	-0.91553	19	-0.26532	31	0.80404	45	1.93022
10	9	-0.91553	22	-0.12479	30	0.66312	42	1.58744
11	6	-1.07021	12	-0.65801	30	0.66312	41	1.42965
12	12	-0.65801	25	-0.02999	37	1.02657	42	1.58744
13	9	-0.91553	18	-0.29783	30	0.66312	45	1.93022
14	9	-0.91553	22	-0.12479	30	0.66312	42	1.58744
15	9	-0.91553	29	0.58512	37	1.02657	45	1.93022
16	6	-1.07021	12	-0.65801	25	-0.02999	41	1.42965
17	10	-0.90007	19	-0.26532	29	0.58512	37	1.02657
18	10	-0.90007	19	-0.26532	29	0.58512	38	1.30327
19	13	-0.64124	25	-0.02999	37	1.02657	44	1.81616
20	9	-0.91553	29	0.58512	38	1.30327	45	1.93022
21	9	-0.91553	19	-0.26532	30	0.66312	45	1.93022
22	9	-0.91553	22	-0.12479	30	0.66312	42	1.58744
23	6	-1.07021	12	-0.65801	26	0.21032	41	1.42965
24	12	-0.65801	25	-0.02999	38	1.30327	44	1.81616
Rasch Parameters								
Mean		-0.912		-0.123		0.747		1.631
SD		0.119		0.395		0.374		0.263
M-1SD		-1.032		-0.517		0.373		1.368
M+1SD		-0.793		0.272		1.120		1.894
Raw Score Parameters								
Mean Cut		13.0		20.0		28.0		36.5
Low Cut		12.0		15.0		20.5		30.5
Cut-1SD		12.0		16.5		24.5		34.0
Cut+1SD		14.0		23.5		31.5		38.5
High Cut		15.0		27.5		33.5		39.0

Using tables 2 and 3, MI staff prepared figures showing the distributions of bookmarks. These distributions are shown in Figures 2 and 3 for Mathematics and Reading, respectively.

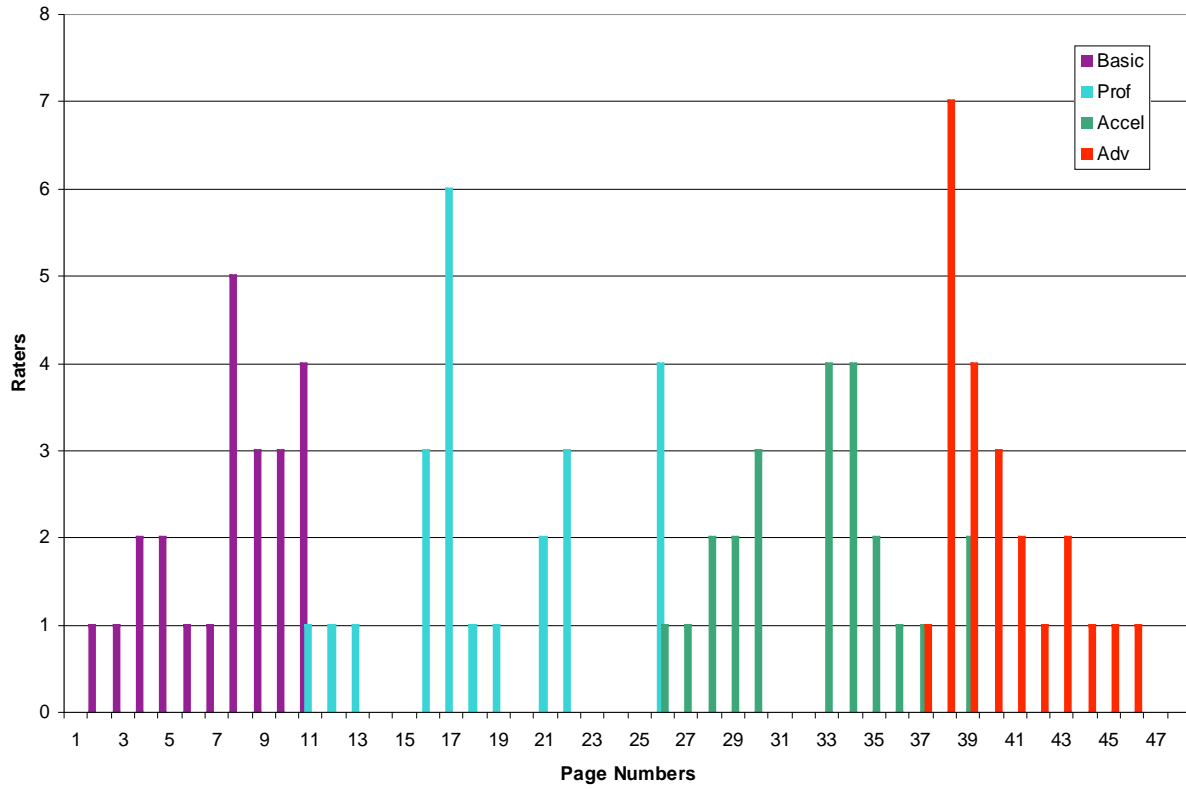


Figure 2. Round 1 distribution of bookmarks for Mathematics

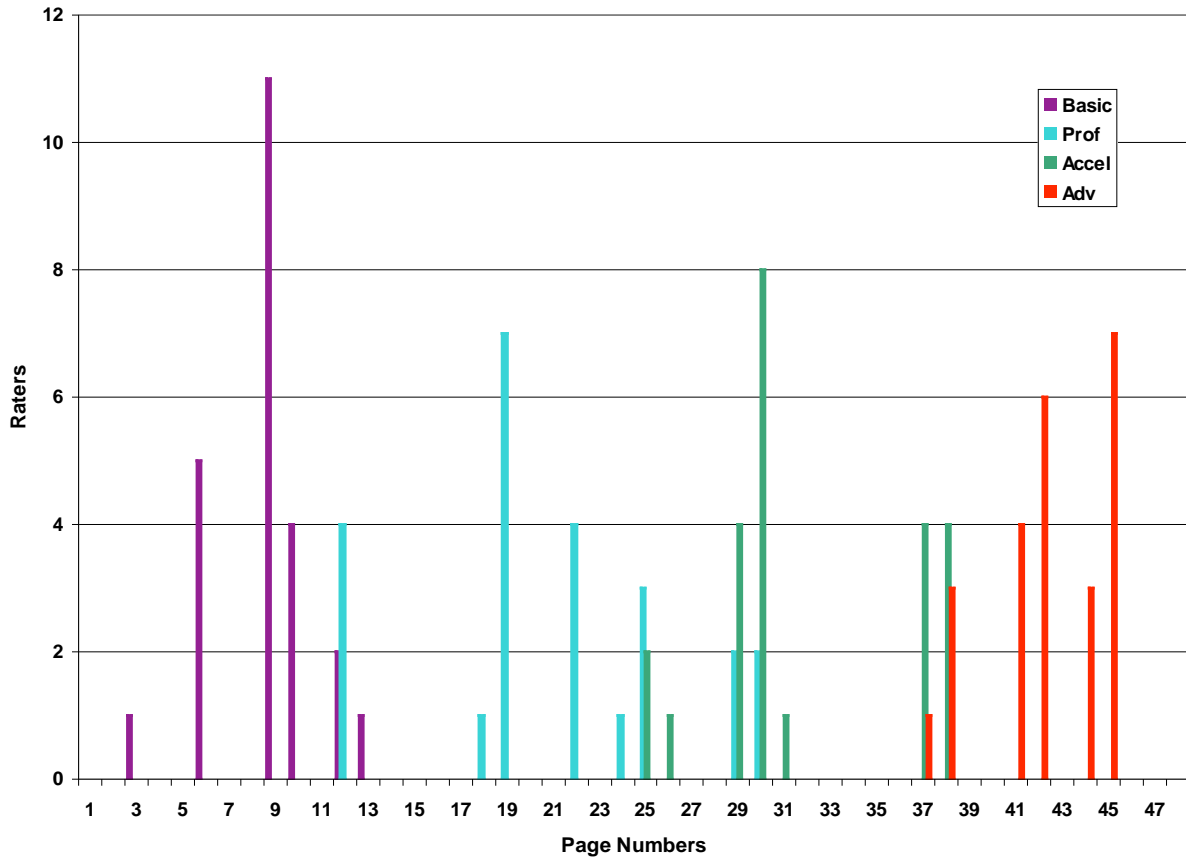


Figure 3. Round 1 distribution of bookmarks for Reading

MI staff had also previously calculated the raw score distributions for both Reading and Mathematics based on performances of students who took the tests in March. They rounded the mean cuts for both tests and applied them to the raw score distributions to create impact data for the entire group of students who took the tests. The same process was also performed by race and gender. All impact data are included in Appendix C.

Round 2

On April 29, Drs. Bunch and Inman convened their respective groups at 8:30 a.m. and distributed the Round 1 bookmarks, test booklets, and other materials including Tables 2 and 3. Large-group discussion of Round 1 began with discussion of Tables 2 and 3 and Figures 2 and 3. Raters discussed the reasons for placing individual bookmarks as they had and what those placements represented. There was then further general discussion of the nature of the just barely Proficient student as well as the just barely Basic student and the just barely Accelerated and barely Advanced student, as well as the demands of the tests as a whole and of individual items. Of particular interest were comments regarding the intersection of the cognitive demands of individual items and the achievements and capabilities of students in each of the categories.

Drs. Bunch and Inman then introduced the impact data in both tabular and graphic form (included in Appendix C) and led discussions of their implications. Results summarized in Table 4 reflect the final tabulations of scores with whole and half-point scores present.

Table 4
Impact Data Summary for Round 1

Category	Mathematics		Reading	
	Cut Score (Out of 46 Points)	Percent in Category	Cut Score (Out of 48 Points)	Percent in Category
Below Basic	---	15.0	---	11.0
Basic	13.0	17.1	13.0	11.4
Proficient	19.0	24.8	20.0	16.5
Accelerated	27.5	14.4	28.0	26.0
Advanced	33.0	28.7	36.5	35.1

Participants in each room focused on three key questions about the Round 1 results:

1. Is the placement of a cut score for Proficient at 41 percent (Mathematics) or 42 percent (Reading) of total possible points realistic?
2. Is it realistic that 68 percent (Mathematics) and 78 percent (Reading) of students would be considered Proficient or above?
3. Is it realistic that the numbers of students in the Advanced category would be greater than the numbers of students in the Proficient or Accelerated categories?

After considerable discussion in each room, raters came to generally positive responses to all three of these questions. They also discussed a variety of procedural issues, the performance level descriptors, and the content of the tests. As discussion drew to a close, Dr. Bunch and Dr. Inman directed participants to respond to the following statement:

I have discussed the results of Round 1, including my ratings, the ratings of others, and the impact data, and I understand what I need to do to complete Round 2.

(Circle one): **No** **Yes**

Raters rejoined their teams and received their bookmarks, ordered test booklets, and other materials needed to complete Round 2. Instructions were the same as for Round 1. Raters were to work alone in silence to place three bookmarks, starting with Proficient. They were to use their first round ratings, the comments they had heard during the morning discussion, the tables and graphs they had received or viewed, and the impact data as they found any one or all pieces of information useful. After all members of a team had entered all three bookmarks, the team could discuss their ratings and make a final adjustment before turning in their individual bookmarks. As individual raters completed their Round 2 tasks, Drs. Inman and Bunch collected their bookmarks, accounted for secure materials, and dismissed them for lunch.

After the last rater in each group had been dismissed for lunch, MI staff entered the page numbers and associated Rasch-based achievement levels in the same Microsoft Excel workbooks they had used for Round 1. Results are summarized in Tables 5 and 6 and Figures 4 and 5. Tables 5 and 6 may be interpreted exactly as Tables 2 and 3 were. The

gray bar at Rater 13 in Table 6 indicates that one rater had to drop out of the group due to a family emergency. Means for Rounds 2 and 3 for Reading were based on 19 raters, rather than 20. Placement of bookmarks by individual raters is shown in Figures 4 and 5.

Table 5
Results for Mathematics Round 2

Rater	Basic Pg	Basic Ach	Prof Pg	Prof Ach	Acc Pg	Acc Ach	Adv Pg	Adv Ach
1	8	-0.6275	17	-0.24375	40	1.05822	46	1.91588
2	10	-0.51333	21	0.05466	39	1.04881	45	1.63433
3	8	-0.6275	17	-0.24375	35	0.83748	45	1.63433
4	11	-0.49392	24	0.324433	39	1.04881	46	1.91588
5	8	-0.6275	18	-0.14562	35	0.83748	44	1.47756
6	8	-0.6275	17	-0.24375	35	0.83748	46	1.91588
7	11	-0.49392	26	0.37107	39	1.04881	45	1.63433
8	8	-0.6275	21	0.05466	39	1.04881	42	1.18172
9	6	-1.08532	16	-0.29154	34	0.8333	46	1.91588
10	11	-0.49392	26	0.37107	40	1.05822	46	1.91588
11	8	-0.6275	17	-0.24375	35	0.83748	45	1.63433
12	8	-0.6275	18	-0.14562	35	0.83748	45	1.63433
13	5	-1.10239	16	-0.29154	28	0.45274	45	1.63433
14	8	-0.6275	17	-0.24375	35	0.83748	45	1.63433
15	8	-0.6275	17	-0.24375	36	0.93394	41	1.11953
16	10	-0.51333	22	0.127362	35	0.83748	44	1.47756
17	10	-0.51333	17	-0.24375	40	1.05822	46	1.91588
18	11	-0.49392	22	0.127362	37	0.99888	45	1.63433
19	9	-0.58382	17	-0.24375	33	0.66094	46	1.91588
20	9	-0.58382	17	-0.24375	35	0.83748	44	1.47756
21	9	-0.58382	21	0.05466	35	0.83748	43	1.43293
22	7	-0.71657	17	-0.24375	35	0.83748	45	1.63433
23	11	-0.49392	21	0.05466	33	0.66094	44	1.47756
Rasch Parameters								
Mean		-0.622		-0.077		0.882		1.642
SD		0.162		0.223		0.154		0.230
M-1SD		-0.784		-0.300		0.727		1.412
M+1SD		-0.460		0.146		1.036		1.872
Raw Score Parameters								
Mean Cut		14.0		19.0		29.5		37.0
Low Cut		10.5		17.0		25.0		32.5
Cut-1SD		13.0		17.0		28.0		35.0
Cut+1SD		15.5		21.5		31.5		39.0
High Cut		15.0		24.0		31.5		39.0

Table 6
Results for Reading Round 2

Rater	Basic Pg	Basic Ach	Prof Pg	Prof Ach	Acc Pg	Acc Ach	Adv Pg	Adv Ach
1	9	-0.91553	22	-0.12479	35	0.93852	42	1.58744
2	9	-0.91553	21	-0.15295	37	1.02657	45	1.93022
3	10	-0.90007	20	-0.17559	38	1.30327	45	1.93022
4	8	-0.92221	18	-0.29783	37	1.02657	41	1.42965
5	10	-0.90007	19	-0.26532	38	1.30327	45	1.93022
6	6	-1.07021	19	-0.26532	37	1.02657	44	1.81616
7	12	-0.65801	28	0.4817	39	1.34542	45	1.93022
8	11	-0.873269	21	-0.15295	38	1.30327	45	1.93022
9	9	-0.91553	20	-0.17559	32	0.8045	45	1.93022
10	11	-0.873269	21	-0.15295	38	1.30327	45	1.93022
11	11	-0.873269	21	-0.15295	38	1.30327	45	1.93022
12	12	-0.65801	25	-0.02999	37	1.02657	42	1.58744
13								
14	9	-0.91553	20	-0.17559	32	0.8045	45	1.93022
15	9	-0.91553	20	-0.17559	35	0.93852	45	1.93022
16	9	-0.91553	20	-0.17559	32	0.8045	45	1.93022
17	11	-0.873269	22	-0.12479	31	0.80404	39	1.34542
18	10	-0.90007	21	-0.15295	37	1.02657	45	1.93022
19	9	-0.91553	21	-0.15295	36	0.9515	45	1.93022
20	9	-0.91553	22	-0.12479	35	0.93852	42	1.58744
21	9	-0.91553	25	-0.02999	31	0.80404	45	1.93022
22	9	-0.91553	22	-0.12479	35	0.93852	42	1.58744
22	9	-0.91553	20	-0.17559	35	0.93852	42	1.58744
24	11	-0.873269	21	-0.15295	38	1.30327	45	1.93022
Rasch Parameters								
Mean		-0.889		-0.132		1.042		1.804
SD		0.083		0.147		0.196		0.194
M-1SD		-0.972		-0.279		0.846		1.610
M+1SD		-0.806		0.015		1.238		1.997
Raw Score Parameters								
Mean Cut		13.0		20.0		31.0		38.0
Low Cut		12.0		18.5		28.5		34.0
Cut-1SD		12.5		18.5		29.0		36.5
Cut+1SD		14.0		21.0		32.5		39.5
High Cut		15.0		25.5		34.0		39.0

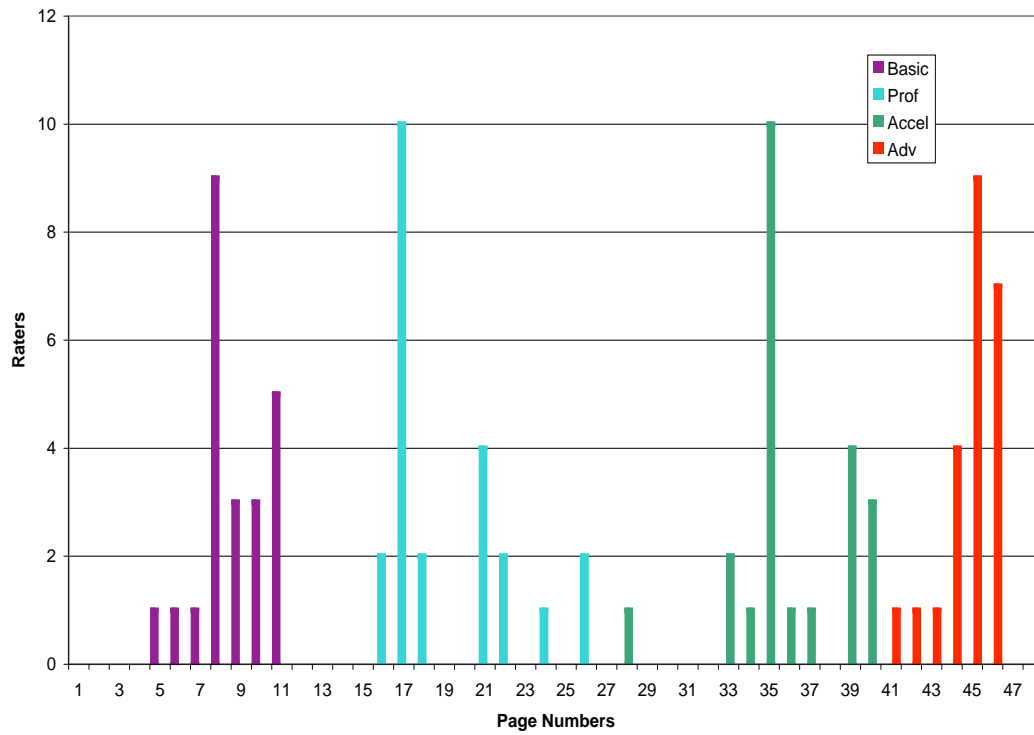


Figure 4. Round 2 distribution of bookmarks for Mathematics

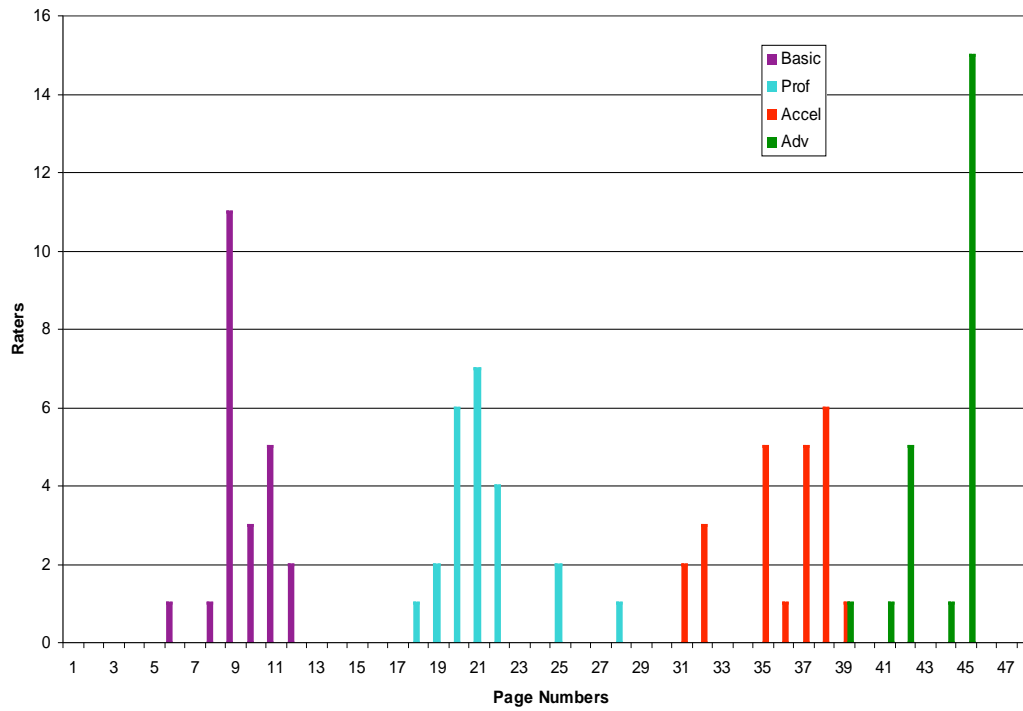


Figure 5. Round 2 distribution of bookmarks for Reading

While there was little movement from Round 1 to Round 2 with regard to overall cut score means, there was some movement from the edges of the distributions toward the means. In Round 2 bookmark placements, there were no overlaps at any level for Mathematics and only one (at score point 39) for Reading. By Round 2, the distributions of the bookmarks were becoming salient.

Round 3

Drs. Inman and Bunch distributed and led discussions of the Round 2 results, pointing out movement in the placement of bookmarks toward the means. Once again, raters discussed their rationales for placing bookmarks as they did, why the percentages of students classified as Below Basic (or limited), Basic, Proficient, Accelerated, or Advanced were appropriate or inappropriate, and the relationship between item cognitive demands and the achievement level definitions. After considerable discussion, raters completed their Readiness Forms, responding to the following statement:

I have discussed the results of Round 2, including my ratings, the ratings of others, and the impact data, and I understand what I need to do to complete Round 3.

(Circle one): **No** **Yes**

Drs. Inman and Bunch inspected the forms and noted that all raters had answered “Yes.” They then directed raters to begin Round 3 following essentially the same directions they had followed in Rounds 1 and 2, with a slight difference. Round 3 ratings were to include the page number and the actual cut scores and associated percentages of students scoring at or above that cut score, according to the tables distributed at the beginning of Round 2. The purpose of this shift in focus was to allow each rater to leave the session knowing at least where he or she had set a recommended cut score and the impact that score would have on students.

As raters completed their Round 3 bookmarks and turned them in, MI staff checked the three sets of numbers to make sure there was no confusion. Every rater gave completely consistent final ratings and cut scores. Raters then completed the remaining portion of the Readiness Form and an evaluation form created specifically for this standard-setting activity. The final four statements on the Readiness Form were as follows:

Round 3: I have completed my ratings, and I believe that the cut scores I have identified fairly represent minimal performances of students at the Basic, Proficient, Accelerated and Advanced levels

(Circle one): **No** **Yes**

Everyone was encouraged to share his or her ratings and hear those of other raters.

(Circle one): **No** **Yes**

The cut scores we recommended accurately reflect the Basic, Proficient, Accelerated, and Advanced achievement levels.

(Circle one): **No** **Yes**

The process was fair and unbiased.

(Circle one): **No** **Yes**

MI staff checked the Readiness Forms as raters turned in their materials prior to being dismissed. There was not a single negative response on any of the 46 Readiness Forms collected.

Ms. Crandell thanked the members of each group for their three days of effort and willingness to be a part of the process by which performance standards on the Ohio Graduation Tests are set. Dr. William Batchelor, MI project director also thanked the groups for their participation and distributed expense checks. Drs. Bunch and Inman collected final ratings and all other secure materials. After each rater's materials were accounted for, he or she was dismissed with a final word of thanks from Dr. Bunch or Dr. Inman.

Final Results

Tables 7 and 8 and Figures 6 and 7 summarize the results of Round 3. In Mathematics, overall cut scores for Basic and Proficient fell by one point each, while the cut score for Advanced remained unchanged from Round 2. In Reading, there was very little movement in mean cut scores, but there was considerable convergence, particularly at the Proficient cut scores. The final mean for the Proficient cut score was 20.0 while the cut scores associated with the mean plus or minus one standard deviation were 18.0 and 21.0, indicating a very pronounced tightening of the range of scores.

Table 7
Results for Mathematics Round 3

Rater	Basic Pg	Basic Ach	Prof Pg	Prof Ach	Acc Pg	Acc Ach	Adv Pg	Adv Ach
1	7	-0.71657	17	-0.24375	40	1.05822	46	1.91588
2	10	-0.51333	19	-0.1201	35	0.83748	44	1.47756
3	8	-0.6275	17	-0.24375	35	0.83748	46	1.91588
4	10	-0.51333	23	0.29697	36	0.93394	46	1.91588
5	8	-0.6275	18	-0.14562	35	0.83748	44	1.47756
6	6	-1.08532	17	-0.24375	40	1.05822	46	1.91588
7	8	-0.6275	26	0.37107	39	1.04881	45	1.63433
8	8	-0.6275	21	0.05466	39	1.04881	42	1.18172
9	6	-1.08532	16	-0.29154	34	0.8333	46	1.91588
10	10	-0.51333	26	0.37107	40	1.05822	46	1.91588
11	8	-0.6275	17	-0.24375	35	0.83748	45	1.63433
12	8	-0.6275	18	-0.14562	37	0.99888	46	1.91588
13	5	-1.10239	16	-0.29154	33	0.66094	45	1.63433
14	8	-0.6275	17	-0.24375	39	1.04881	45	1.63433
15	8	-0.6275	17	-0.24375	36	0.93394	43	1.43293
16	10	-0.51333	26	0.37107	37	0.99888	46	1.91588
17	8	-0.6275	17	-0.24375	40	1.05822	46	1.91588
18	8	-0.6275	26	0.37107	37	0.99888	45	1.63433
19	8	-0.6275	17	-0.24375	35	0.83748	46	1.91588
20	8	-0.6275	17	-0.24375	35	0.83748	44	1.47756
21	8	-0.6275	22	0.127362	35	0.83748	43	1.43293
22	7	-0.71657	17	-0.24375	35	0.83748	45	1.63433
23	11	-0.49392	19	-0.1201	35	0.83748	44	1.47756
Rasch Parameters								
Mean		-0.670		-0.069		0.925		1.692
SD		0.177		0.251		0.113		0.224
M-1SD		-0.847		-0.320		0.812		1.469
M+1SD		-0.493		0.182		1.038		1.916
Raw Score Parameters								
Mean Cut		13.5		19.5		30.0		37.5
Low Cut		10.5		17.0		27.0		33.0
Cut-1SD		12.5		17.0		29.0		35.5
Cut+1SD		15.0		22.0		31.5		39.5
High Cut		15.0		24.0		31.5		39.0

Table 8
Results for Reading Round 3

Rater	Basic Pg	Basic Ach	Prof Pg	Prof Ach	Acc Pg	Acc Ach	Adv Pg	Adv Ach
1	10	-0.90007	20	-0.17559	35	0.93852	42	1.58744
2	9	-0.91553	21	-0.15295	38	1.30327	46	2.19568
3	10	-0.90007	20	-0.17559	38	1.30327	46	2.19568
4	8	-0.92221	17	-0.34661	37	1.02657	41	1.42965
5	10	-0.90007	20	-0.17559	38	1.30327	46	2.19568
6	8	-0.92221	19	-0.26532	38	1.30327	46	2.19568
7	9	-0.91553	27	0.31562	41	1.42965	46	2.19568
8	13	-0.64124	21	-0.15295	38	1.30327	45	1.93022
9	10	-0.90007	22	-0.12479	35	0.93852	45	1.93022
10	13	-0.64124	21	-0.15295	38	1.30327	45	1.93022
11	13	-0.64124	21	-0.15295	38	1.30327	45	1.93022
12	13	-0.64124	25	-0.02999	39	1.34542	44	1.81616
14	9	-0.91553	22	-0.12479	37	1.02657	45	1.93022
15	9	-0.91553	22	-0.12479	37	1.02657	45	1.93022
16	9	-0.91553	22	-0.12479	37	1.02657	45	1.93022
17	11	-0.873269	24	-0.06231	35	0.93852	45	1.93022
18	10	-0.90007	21	-0.15295	37	1.02657	45	1.93022
19	9	-0.91553	21	-0.15295	37	1.02657	45	1.93022
20	10	-0.90007	22	-0.12479	35	0.93852	42	1.58744
21	9	-0.91553	19	-0.26532	33	0.807	45	1.93022
22	10	-0.90007	22	-0.12479	35	0.93852	42	1.58744
23	10	-0.90007	22	-0.12479	35	0.93852	42	1.58744
24	13	-0.64124	21	-0.15295	38	1.30327	45	1.93022
Rasch Parameters								
Mean		-0.849		-0.136		1.122		1.902
SD		0.113		0.118		0.186		0.220
M-1SD		-0.962		-0.254		0.935		1.681
M+1SD		-0.737		-0.018		1.308		2.122
Raw Score Parameters								
Mean Cut		13.5		20.0		31.5		39.0
Low Cut		13.0		18.0		28.5		34.5
Cut-1SD		12.5		18.5		30.0		37.0
Cut+1SD		14.5		21.0		33.5		40.5
High Cut		15.5		24.0		34.5		41.0

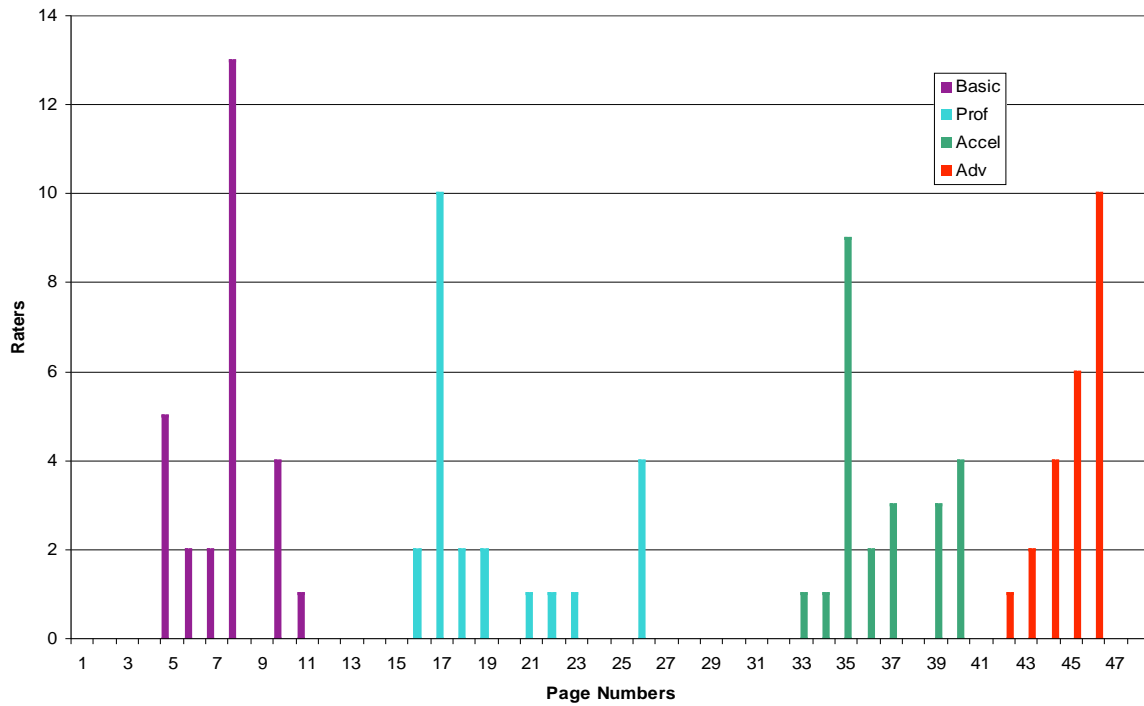


Figure 6. Round 3 distribution of bookmarks for Mathematics

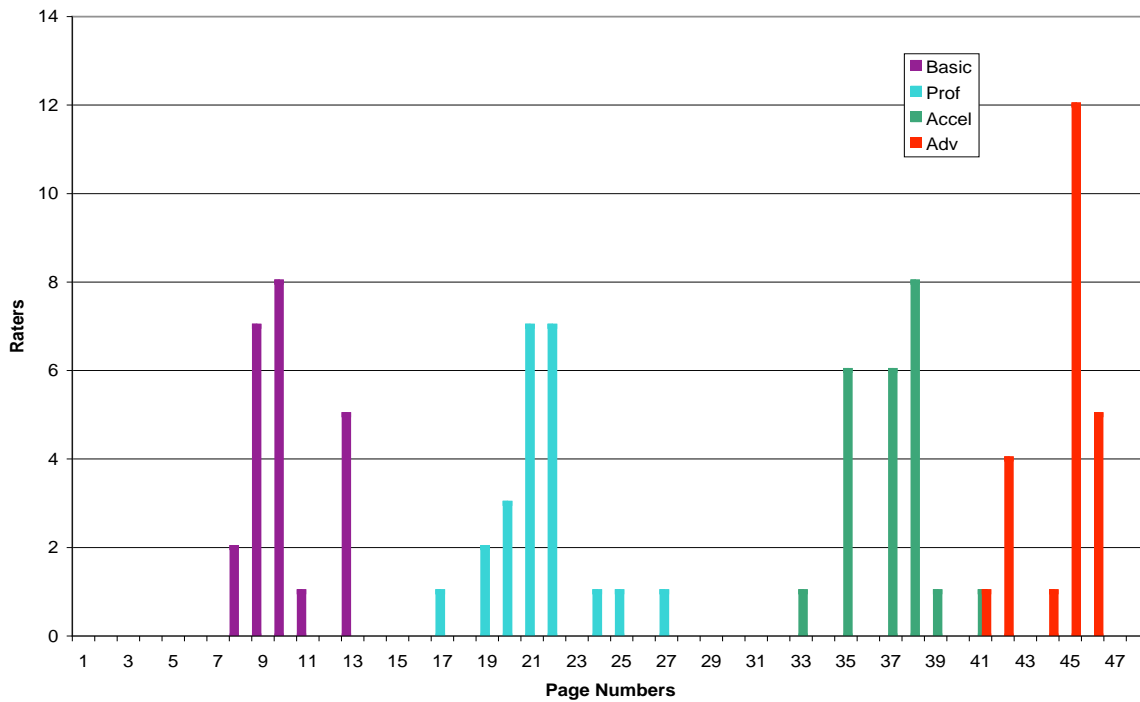


Figure 7. Round 3 distribution of bookmarks for Reading

Tables 9 and 10 summarize the three rounds of ratings and associated raw cut scores over three rounds of ratings. The cut score for Proficient did not change in either subject over three rounds. However, the percentages of students in this category did change because the boundaries for the Accelerated and Advanced categories changed, always upward. Cut scores for Basic changed very little from round to round.

Impact.

Tables 9 and 10 show the final (Round 3) cut scores and impact (percentage of students in each category). Figures 8 and 9 summarize the same information graphically.

Table 9
Cut Scores by Round: Mathematics
(Percentages of students in each group are shown in parentheses.)

Category	Round		
	1	2	3
Limited	--- (15.0)	--- (17.8)	--- (17.2)
Basic	13.0 (17.1)	14.0 (14.3)	13.5 (14.9)
Proficient (Graduation)	19.0 (24.8)	19.0 (30.1)	19.0 (31.3)
Accelerated	27.5 (14.4)	29.5 (19.2)	30.0 (19.6)
Advanced	33.0 (28.7)	37.0 (18.6)	37.5 (17.0)
Percent of Students Proficient or Above	67.9	67.9	67.9

Table 10
Cut Scores by Round: Reading
(Percentages of students in each group are shown in parentheses.)

Category	Round		
	1	2	3
Limited	--- (11.0)	--- (11.3)	--- (12.1)
Basic	13.0 (11.4)	13.0 (11.4)	13.5 (10.3)
Proficient (Graduation)	20.0 (16.5)	20.0 (24.4)	20.0 (25.9)
Accelerated	28.0 (26.0)	31.0 (23.6)	31.5 (25.9)
Advanced	36.5 (35.1)	38.0 (29.6)	39.0 (25.9)
Percent of Students Proficient or Above	77.6	77.6	77.7

**Reading Raw Score Distribution With Round 3 Cuts
and Percentages of Students in Each Category (n=74,337)**

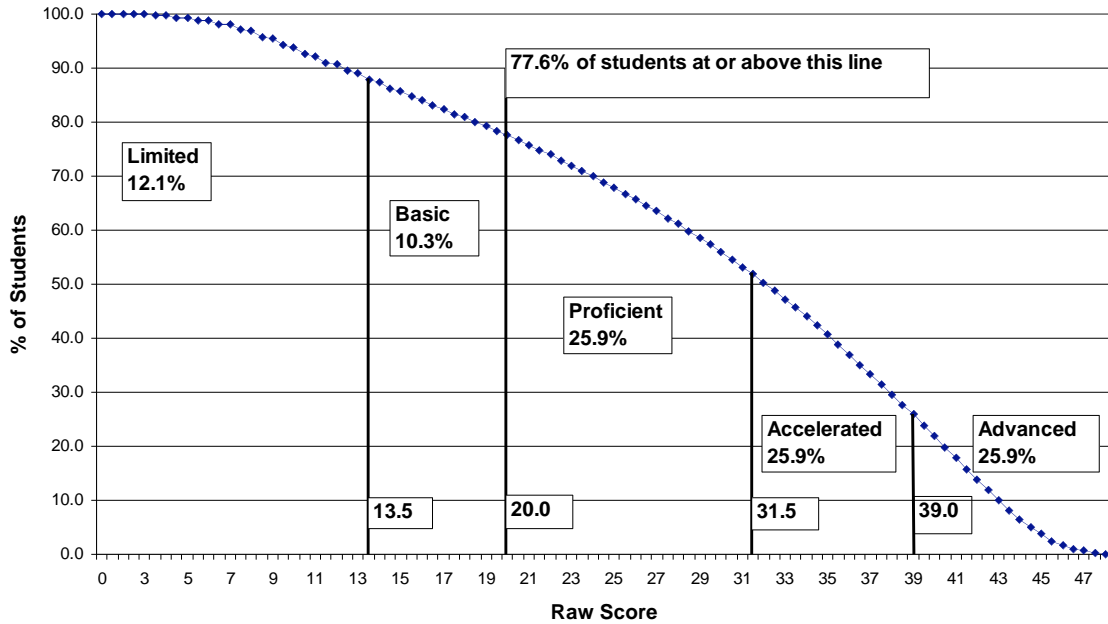


Figure 8. Raw score distribution for Mathematics

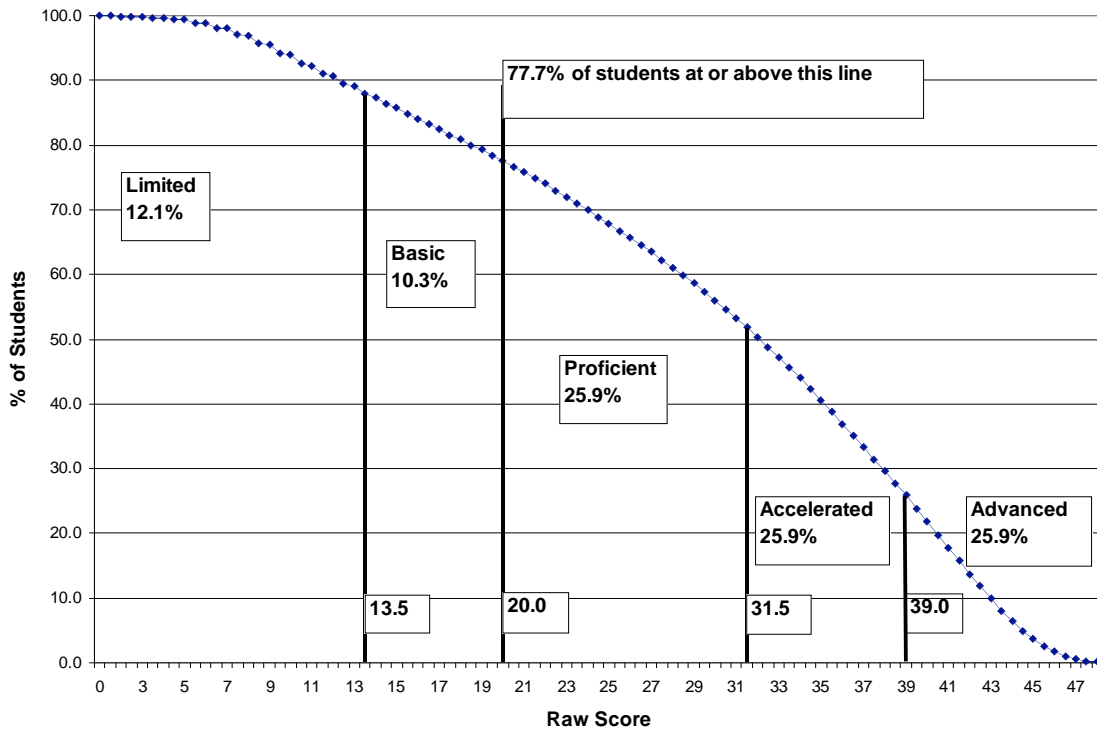


Figure 9. Raw score distribution for Reading

While Tables 9 and 10 and Figures 8 and 9 show the overall distributions of students by achievement level, it does not address possible differences among groups. Tables 12 and 13 show these same distributions by race and sex. For purposes of comparison, the numbers of students in each group were as follows:

Table 11. Examinee counts by group

Mathematics		Reading	
Group	Count	Group	Count
Native American/ American Indian	139	Native American/ American Indian	118
Asian/Pacific Islander	881	Asian/Pacific Islander	871
Black/African American	10,239	Black/African American	10,189
Hispanic	1,499	Hispanic	1,508
White	58,986	White	57,318
Multiracial	940	Multiracial	964
Other Races	204	Other Races	197
	Sum	Sum	71,165
Female	36,395	Female	36,470
Male	38,487	Male	37,678
	Sum	Sum	74,148

Notes: Some examinees failed to report either or both race and gender. The total sample was 74,337 for Reading and 76,092 for Mathematics.

Table 12
Distribution of Students by Race and Sex: Mathematics
(Entries are percentages.)

Group (Tested)	Category					Proficient or Above
	Limited	Basic	Proficient	Accelerated	Advanced	
AmInd (139)	28.78	10.79	34.53	12.95	12.95	56.95
As-PI (881)	8.06	10.22	24.29	21.45	35.98	82.50
BL-AA (10,239)	36.43	24.65	28.59	7.82	2.51	40.08
Hisp (1,499)	30.82	22.08	29.62	10.81	6.67	45.16
Multi (940)	23.40	16.49	30.43	16.49	13.19	60.11
Other (204)	28.92	15.20	35.29	7.84	12.75	55.88
White (58,986)	13.20	12.93	31.95	22.19	19.72	73.86
Total by Race 72,888	16.97	14.78	31.33	19.80	17.12	68.25
Female (37,395)	14.40	15.60	34.04	20.19	15.77	70.00
Male (38,487)	19.86	14.17	28.61	19.14	18.22	65.97
Total by Sex 75,882	17.17	14.87	31.29	19.66	17.01	67.96

Table 13
Distribution of Students by Race and Sex: Reading
(Entries are percentages.)

Group (Tested)	Category					Proficient or Above
	Limited	Basic	Proficient	Accelerated	Advanced	
AmInd (118)	21.19	16.95	26.27	24.58	11.02	61.87
As-PI (871)	8.61	7.12	20.67	25.49	38.12	84.28
BL-AA (10,189)	23.88	17.73	33.07	17.35	7.96	58.38
Hisp (1,508)	22.08	18.30	28.12	20.29	11.21	59.62
Multi (964)	14.73	13.07	27.07	23.55	21.58	72.2
Other (197)	22.84	12.69	29.44	21.32	13.71	64.47
White (57,318)	9.38	8.64	24.43	27.74	29.81	81.98
Total by Race 71,165	11.84	10.22	25.75	25.99	26.20	77.94
Female (36,470)	7.28	8.68	26.32	27.37	30.35	84.04
Male (37,678)	16.58	11.90	25.41	24.48	21.63	71.52
Total by Sex 74,148	12.01	10.32	25.86	25.90	25.92	77.68

Readiness and evaluations.

As noted above, there were no negative responses to any of the statements on the Readiness Form. Two raters in the Reading group asked about the statement regarding the cut scores accurately reflecting the three achievement levels. Because they did not yet know the results of Round 3, they left this item blank. They both noted, however, that the results from Rounds 1 and 2 accurately reflected the three achievement levels.

Results of the evaluation were extremely positive, as shown in Table 14. In the Reading group, evaluations were unanimously positive – 100 percent agreement with all statements. In the Mathematics group, there were two negative responses out of 189; one rater did not think the impact data were helpful, and one rater thought the room was too small. Otherwise, all raters in both groups gave overwhelmingly positive evaluations.

Table 14
Summary of Evaluations
(Entries are percentages.)

Statement		Mathematics		Reading	
		Agree	Disagree	Agree	Disagree
1	The orientation provided me with a clear understanding of the purpose of the meeting.	96	4	100	0
2	The workshop leaders clearly explained the task.	96	4	100	0
3	The training and practice exercises helped me understand how to perform the task.	100	0	100	0
4	Taking the test helped me to understand the assessment.	100	0	100	0
5	The performance level descriptions were clear and useful.	100	0	86	14
6	The large and small group discussions aided my understanding of the process.	100	0	100	0
7	The time provided for discussions was adequate.	96	4	100	0
8	There was an equal opportunity for everyone in my group to contribute his/her ideas and opinions.	100	0	100	0
9	I was able to follow the instructions and complete the rating sheets accurately.	100	0	100	0
10	The discussions after the first round of rating were helpful to me.	100	0	100	0
11	The discussions after the second round of rating were helpful to me	100	0	100	0
12	The information showing the distribution of student scores was helpful to me.	100	0	100	0
13	I am confident about the defensibility and appropriateness of the final recommended cut scores.	100	0	100	0
14	The facilities and food service helped create a productive and efficient working environment.	96	4	95	5

Comments.

Members of both groups gave overwhelmingly positive comments about the process and outcomes. The only negative comments tended to be about the physical setting (temperature and light) and hotel accommodations (the hotel was being renovated at the time). All ratings and comments are included in Appendix D.

Follow-up Activities

In accordance with the plan for standard setting, ODE and MI staff presented the results and background information to three groups with acute interest in the tests: the Technical Advisory Committee, the Test Steering Committee, and the Fairness and Sensitivity Review Committee. The reactions of these groups are presented below.

TAC.

On the morning of May 4, 2004, MI staff, ODE staff, and nine members of the TAC held a telephone conference to discuss the standard setting. Conclusions and recommendations are summarized below:

- a. It was recommended that an introductory paragraph be added to the report.
- b. TAC members generally agreed that it was not a good practice to sort the results of a 38-item test, such as the Ohio Graduation Tests, into 5 performance categories but under the circumstances the results were deemed to be very good.
- c. It also was agreed that the report should focus on the product of standard setting and not the process.
- d. The recommended standards resulted in percent proficient approximately the same as the original standards for the Ohio Ninth Grade Proficiency Tests when those tests were first adopted.
- e. Discussion supported the notion that the conjunctive nature of the model used for standard setting was an important feature.
- f. Emphasis on the Proficient Level throughout the report was recommended.
- g. It was pointed out that summing categories of data within some of the tables did not produce a total equal to the total involved in the sample. There was general agreement that an explanation of this anomaly should be included in the final draft of the report.
- h. The impact of the recommended cut scores on the largest student demographic groups was underscored as an important feature of the report.
- i. It was recommended that percentage of students affected by the recommended cut scores be included in Table ES-5.
- j. The group agreed that the presentation to the Test Steering Committee should focus on impact of the recommended cut scores on the overall student population and not necessarily involve discussion of details such as 1/2 point issues at various cut levels.

TSC.

On May 5, ODE and MI staff presented results and background information to the Test Steering Committee. While much of this group's discussion focused on mechanics of the tests and administration, there was considerable interest in the results for the Mathematics test. In particular, this group was gratified with the results of the 2004 test administration. They supported the process and the outcomes.

Conclusions

Standard setting is a combination of art and science. It combines the democratic process of group interaction and decision making with carefully planned and executed steps based on a well-defined mathematical model. The process by which the standard-setting activity for the Reading and Mathematics tests of the OGT was carried out, described in detail in the body of this report, was meticulously crafted by experienced psychometricians and reviewed by a national body of experts in the field. The plan was carried out under the supervision of ODE staff and an external reviewer who is also an expert in this field. Because there are no “true” cut scores for any test, the recommended cut scores are only as valid as the process by which they were derived.

The people who worked through the process understood what was expected of them, and they remained on task throughout the three days. Round by round, they shared their rationales and heard one another out. In the end, there were still disagreements about exactly where cut scores should be, but the disagreements were not divisive, and there was general consensus that the results were reasonable and fair. The groups who reviewed the results – the TAC and TSC – expressed their support for the process and results. By all accounts, this standard-setting activity was well planned and executed. The recommended cut scores should be taken as valid.

Appendix A: Standard Setting Plan

Spring 2004

Plan updates

Draft Plan (November 2003)

Plan Updates

The original plan, drafted in November 2003, was based on the assumption of four performance levels: Limited, Basic, Proficient, and Advanced. House Bill 3 (HB3) added a fifth category, Accelerated, between Proficient and Advanced. The addition of this new level required minor modification of the procedures and forms used during the process.

At the February 2004 meeting of the Technical Advisory Committee, the issue of probability of correct response arose. The TAC recommended that the fundamental standard-setting question be changed from a focus on a two-thirds chance of answering an item correctly to a fifty-percent chance of answering correctly. This change in orientation of the participants prompted some modification of the overview presentations. The major modification to the plan, however, was the elimination of a step that is described in Attachment C (Procedures For Translating Rasch Step Values Into Achievement Levels) of this Plan. Because it was no longer necessary to calculate the Rasch ability level associated with a two-thirds chance of answering correctly, we were able simply to use the Rasch threshold values calculated by WINSTEPS or BIGSTEPS. We had originally planned to use these values as check values for our calculations of the Rasch ability levels associated with a two-thirds chance of answering correctly. Winsteps produces the 50 percent probability Rasch ability level as a byproduct for both dichotomous and polytomous items.

In the training and execution stages of standard setting, the fundamental standard setting instructions and question changed only slightly. The original instructions included the following:

1. What knowledge or skill is required to earn this point?
2. What makes later items more difficult than earlier items?
3. Think of a large group of students at the cut score for this level. Would about 2/3 of them earn this point?

In the final execution of the plan, only question 3 was changed: “2/3” was changed to “50 percent.” Forms used by participants to enter their ratings were modified to reflect this change.

The change from two-thirds to fifty percent also affected the production of difficulty-ordered booklets. Item ordering for a two-thirds probability of correct response will not necessarily be the same as ordering when the probability is fifty percent.

There were no other major changes.

Ohio Graduation Tests Standard Setting Plan: Spring 2004

November 2003

Michael B. Bunch

Measurement Incorporated

In April 2003, Measurement Incorporated (MI), under contract to the Ohio Department of Education (ODE) led standard-setting activities in Columbus for the Reading and Mathematics tests that had been administered the previous month. The plan for conducting those standard-setting activities had been previously approved by the ODE Technical Advisory Committee (TAC) and carried out by MI staff. The TAC subsequently reviewed the results of the standard-setting activities and found the entire procedure to have been technically sound. Results were forwarded to the Ohio Board of Education (Board), which did not approve a standard for either test. Thus a second standard-setting activity has been planned for both tests on April 27 - 29, 2004.

This plan, based on the plan previously approved by the TAC and carried out in 2003, focuses specifically on activities to take place on April 27-29, 2004 with the Reading and Mathematics portions of the OGT. The persons involved, the data required, and the seven steps we will take to complete the task are described in detail in this Plan:

- Public Engagement
- Preparation
- Training
- Test Administration
- Practice Test
- Rating Test Items
- Follow-Up and Reporting

Drs. Michael Bunch and Elliot Inman will conduct the standard-setting meetings, as they did in 2003. Both are experienced standard-setting meeting facilitators and are well known to ODE staff. We propose to invite 25 participants (Ohio educators, parents, and community leaders) per content area to standard-setting sessions to be held in the spring of 2004 for Reading and Mathematics.

We propose a three-day standard setting session for each test, employing a Rasch-based modified bookmark procedure (Mitzel, Lewis, Patz, & Green, 2001). Prior to the standard-setting meetings, MI will analyze data from the operational tests and calculate Rasch statistics for each item. In preparation for the standard setting meeting, we will rank order the items by IRT difficulty and prepare difficulty-ordered booklets as well as participant training materials including overhead transparencies, guides, and forms.

Public Engagement. We will share final versions of achievement level descriptors with the Test Steering Committee (TSC) and the Fairness and Sensitivity Review Committee (FSRC). We will work directly with ODE staff in the development of achievement level descriptors. The ODE has requested that we share the plans and results of the standard setting activities with the TSC and FSRC prior to the presentation of recommendations to the State Board of Education in May 2004. We will do so and incorporate their comments in our report to the Board.

Preparation. We propose to use a Rasch-based modified bookmark procedure for standard setting as we did in 2003. This procedure features a difficulty-ordered test booklet which raters use to indicate points at which students of a specified achievement level would no longer have a reasonable chance of answering items correctly. Raters mark these points with “bookmarks” to indicate cut points between students of one level of achievement and those of the next higher level.

Participants. While selection of method and facilitators is important to the success of any standard setting activity, the selection and orientation of standard-setting participants is crucial. The ODE has taken responsibility for identifying, recruiting, and securing the services of standard setters to whom we will refer as “raters” to distinguish their advisory role from the policy-making role of the State Board of Education. MI will provide whatever assistance the ODE requests and will follow up on leads initiated by the ODE.

It is likely that the pool of raters will include classroom teachers and building-level administrators as well as non-educators (e.g., parents, community leaders). We also anticipate that some teachers will cross subjects; i.e., some reading specialists will rate mathematics tests, and some mathematics teachers will rate reading tests. We will maintain a list of participants and their relevant characteristics (gender, race, educator/non-educator, subject taught) for subsequent analysis of results by group, but we will not directly associate a participant’s name with that participant’s ratings.

Materials. MI staff will prepare materials for the spring 2004 OGT standard setting. ODE staff will select participants. Except for participant selection, MI will be responsible for all aspects of standard setting. Materials will include the following:

Follow-up letters with instructions

- PowerPoint presentation slides
- PowerPoint presentation participant handouts
- PowerPoint presenter's notes
- Practice tests and scoring guides
- Achievement level descriptors
- Difficulty-ordered test booklets and scoring guides
- Bookmarks Item data sheets
- Impact data sheets
- Data entry and processing programs and procedures
- Data presentation spreadsheets and graphic displays of rating data
- Readiness Form Evaluation forms

Attachment A of this Plan contains sample materials.

Achievement levels. House Bill 3 (HB3) calls for a new set of performance levels with four rather than three cut scores and five rather than four categories: *Limited*, *Basic*, *Proficient*, *Accelerated*, and *Advanced*. MI staff will work with ODE staff to draft working definitions of all five levels. We will draw upon existing level descriptors (e.g., those in 2003 for four achievement levels) and adapt them to the current curriculum. In preparing the drafts, MI staff will follow an iterative process with ODE staff until all are satisfied that a suitable draft is ready. We will then work with raters to help them internalize the definitions and apply them consistently. We will give raters a final opportunity to review the definitions after the last round of rating to make sure they continue to be aligned with the recommended cut scores.

The bookmark method. We have proposed to use a modified bookmark method (Mitzel, Lewis, Patz, & Green, 2001). A brief overview of the method is provided here, with emphasis on technical details raised by the TAC in the fall of 2002 and spring of 2003.

The bookmark procedure is so named because raters identify cut scores by entering markers in a specially designed test booklet. The test booklet consists of a set of items placed in difficulty order, easiest items first and hardest items last. In between, multiple choice (MC) and constructed response (CR) items are intermingled in order of their difficulty. Each CR item appears several times in the booklet, once for each of its score points. For a given CR entry, the item prompt and the rubric for a particular score point appear, along with sample responses illustrating that score point. The method has become quite popular because of its ability to present MC and CR items at the same time and because of its use of item response theory (IRT) analyses.

The difficulty-ordered booklet can be composed of any collection of items spanning the range of content, item types, and difficulty represented in a typical test, and

need not consist only of items that have appeared in an intact test. This booklet can have more items or fewer items than a regular test booklet. For the spring 2004 standard-setting activity, we plan to use only the items that appear in the spring 2004 operational test.

With the bookmark procedure, IRT difficulty indices of the MC items and step values of the various score points for the CR items are ordered from least to most difficult. For the Ohio Graduation Test (OGT), we will use a one parameter IRT model, specifically the Rasch model. Each MC item will have one associated Rasch difficulty index, and each CR item will have as many Rasch step (difficulty) functions as it has score points (excluding zero).

Raters will work in small groups, evaluating the contents of small clusters of items as they appear in the difficulty-ordered test booklet. They will discuss what makes one item or group of items more difficult than those that preceded it and ultimately place a bookmark at a point where they believe the difficulty of the subsequent items exceeds the ability of an identified group of students. Thus, for example, raters would begin with the first item and ask themselves if a minimally qualified student (or group of students) at a particular achievement level (e.g., just barely *Proficient* or just barely *Advanced*) would have a reasonable chance of answering the item correctly (cf. Mitzel, *et al.*, 2001, p. 260). They would then ask themselves the same question for each subsequent item until they reached one where they could not answer affirmatively. The final item yielding an affirmative response would mark the beginning of that performance level, and the raters would place a bookmark at that point (i.e., after the last attainable item).

Operationally defining the decision rule. Crucial to the decision about who does and who does not have a reasonable chance of success is the definition of reasonable chance. Typically, bookmark and other item mapping procedures employ a 2/3 rule; i.e., raters are asked to consider whether about two-thirds of the students who are just barely *Proficient* (or *Advanced* or *Basic*) would probably answer the item correctly or earn the score point in question. This operational definition of reasonable chance is central to the calculations necessary to determine cut scores.

We begin with the fundamental equation of the Rasch model for dichotomous items (e.g., Wright & Stone, 1979; equation 1.4.1, p. 15):

$$P(X = 1 | \beta_v, \delta_i) = \frac{\exp(\beta_v - \delta_i)}{[1 + \exp(\beta_v - \delta_i)]} \quad (1)$$

where β_v is the Rasch ability (achievement level) of a student at the cut score (SACS), δ_i is the Rasch difficulty of item i , and \exp is the base of the natural logarithm raised to the power inside the parentheses. Allowing the expression on the right to equal 2/3 and solving for β_v , we obtain the following:

$$\begin{aligned}
\exp(\beta_v - \delta_i) / [1 + \exp(\beta_v - \delta_i)] &= 2/3 \\
\exp(\beta_v - \delta_i) &= [1 + \exp(\beta_v - \delta_i)] * 2/3 \\
\exp(\beta_v - \delta_i) &= 2/3 + 2/3 * \exp(\beta_v - \delta_i) \\
\exp(\beta_v - \delta_i) - 2/3 * \exp(\beta_v - \delta_i) &= 2/3 \\
1/3 * \exp(\beta_v - \delta_i) &= 2/3 \\
\exp(\beta_v - \delta_i) &= 2/3 / 1/3 \\
\exp(\beta_v - \delta_i) &= 2 \\
\beta_v - \delta_i &= \ln(2) \\
-\delta_i &= .693 \\
\beta_v &= \delta_i + .693
\end{aligned}$$

Thus, in order for a student to have a 2/3 chance of answering a given MC item correctly, that student has to have Rasch ability .693 logits greater than the Rasch difficulty of the item.

The OGT also contains several constructed-response (CR) items, all scored on a two-point or four-point scale. For CR items, we employ the Partial-Credit Model (PCM; Wright & Masters, 1982). For CR items, the likelihood (π_{nix}) of a person with a given ability (β_n) obtaining any given score (j) in any item (i) is shown in equation 2, taken from Wright & Masters (1982), equation 3.1.6:

$$\pi_{nix} = \frac{\exp \sum_{j=0}^x (\beta_n - \delta_{ij})}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^k (\beta_n - \delta_{ij})} \quad (2)$$

In Wright & Masters' formulation, the difficulties associated with each score point are referred to as step functions (δ_{ij}). A key concept is that the step function for score point 0 is set equal to 0 for equation (2):

$$\delta_{i0} \equiv 0, \text{ such that } \sum_{j=0}^0 (\beta_n - \delta_{ij}) = 0, \text{ and } \exp \sum_{j=0}^0 (\beta_n - \delta_{ij}) = 1$$

The numerator values for the other steps are derived as follows:

$$\text{Step 1. } \sum_{j=0}^1 (\beta_n - \delta_{ij}) = \sum_{j=0}^0 (\beta_n - \delta_{ij}) + \beta_n - \delta_{i1} = 0 + \beta_n - \delta_{i1} = \beta_n - \delta_{i1} \quad (3)$$

$$\text{Step 2. By similar logic: } \sum_{j=0}^2 (\beta_n - \delta_{ij}) = 2\beta_n - \delta_{i1} - \delta_{i2} \quad (4)$$

Step 3. By similar logic: $\sum_{j=0}^3 (\beta_n - \delta_{ij}) = 3\beta_n - \delta_{i1} - \delta_{i2} - \delta_{i3}$ (5)

Step 4. By similar logic: $\sum_{j=0}^4 (\beta_n - \delta_{ij}) = 4\beta_n - \delta_{i1} - \delta_{i2} - \delta_{i3} - \delta_{i4}$ (6)

The exponential values of these sums are simply the natural logarithm e raised to the respective values:

Step 1. $\exp(\beta_n - \delta_{i1})$;

Step 2. $\exp(2\beta_n - \delta_{i1} - \delta_{i2})$;

Step 3. $\exp(3\beta_n - \delta_{i1} - \delta_{i2} - \delta_{i3})$;

Step 4. $\exp(4\beta_n - \delta_{i1} - \delta_{i2} - \delta_{i3} - \delta_{i4})$.

The denominator of equation (2) now becomes the simple sum of the four values shown above (for a 4-point item) plus 1 (the exponential value for step 0). The likelihood of obtaining any given score (0 through 4) can be calculated by dividing the numerator associated with that score point by this common denominator.

The derivation of the ability necessary to obtain a given score point is not so direct and neat as for multiple-choice (MC) items. Indeed, it is necessary to calculate a system of probabilities for each CR item (i.e., a probability for each score point) simultaneously. It should also be evident from the development above that the probabilities for each score point are interdependent; i.e., the likelihood of obtaining score point 4 must take into account the likelihood of obtaining score points 3, 2, 1, and 0. Rather than derive a general formula for obtaining these likelihoods, we create item characteristic curves and tables (somewhat similar to Reckase charts) and show the calculated probability of obtaining each score point for a wide range of abilities (e.g., -4 to +4 by increments of 0.1). We then look up the desired value. When necessary, we interpolate to obtain exact values or enter additional ability estimates in the first column.

In calculating and plotting these probabilities, it becomes very clear that for some score points, there is no ability value associated with a 2/3 (.67) probability of obtaining that score. Indeed, only 0 and 4 (in the present case, or 0 and any perfect score in general) are likely to yield a .67 probability. Our response to this situation has been to consider not just the likelihood of obtaining a given score but of obtaining that score or better. Thus, for example, we concern ourselves not with the likelihood of obtaining a score of 1 or 2 but of obtaining a score of 1 or better and 2 or better. BIGSTEPS/WINSTEPS calculates the expected score point for $P = 0.5$ (the Thurstone Expected Score Measures) for score point 1 or better, 2 or better, etc. We simply extend the logic of the Thurstone measures to derive an expected

score measure (ability) at $P=.67$. Attachment C contains a more detailed description of the process.

Preparing the difficulty-ordered booklet. As noted above, MI will prepare a difficulty-ordered test booklet consisting of the items in the spring 2004 test booklets. Table A1 shows sample values for the 2003 OGT Reading test. The CR items are listed two or four times each, once for each non-zero score point. This table shows the Rasch difficulty indices (RASCH) for each item, along with item p values. For CR items, the step values for each score point and corresponding percentages of students scoring at each point or better are shown. For example, on page 10 (15-1), the Rasch Achievement value of .27 indicates that students with a Rasch achievement level (θ_n in the formulation above) of .27 would have a 2/3 chance of obtaining a score of 1 or better on item 15.

In Table A1, Column 1 (Item) shows the order of the item in the regular, spring 2003 test booklet. Rasch Difficulty (column 2) is the calibrated value of difficulty for each MC item (δ_i) or step function for each CR item score point (δ_{ik}). The third column shows the achievement level (Rasch ability) required to give a student a 2/3 chance of answering the MC item correctly or achieve that score point or better on a CR item (Ach. @ 2/3 Pass). This final column refers to the probability of passing the item or score point, not the probability of passing the test. A slightly different version of Table A1 will be presented after Round 2. That version will have the cut score associated with the Ach. @ 2/3 Pass.

Table A1 - Ordered Booklet Spreadsheet: Reading 2003

Item	Rasch Difficulty	Ach @ 2/3 Pass	Page
6	-2.305	-1.612	1
2	-1.986	-1.293	2
3	-1.95	-1.257	3
12	-1.304	-0.611	4
14	-1.027	-0.334	5
11	-0.936	-0.243	6
1	-0.869	-0.176	7
28	-0.611	0.082	8
24	-0.5	0.193	9
15-1	0.48	0.27	10
18	-0.421	0.272	11
17	-0.407	0.286	12
26-1	0.79	0.42	13
5-1	0.65	0.44	14
8-1	0.35	0.44	15
7	-0.2	0.493	16
29-1	0.24	0.54	17
27	-0.142	0.551	18
34	-0.135	0.558	19
20	-0.124	0.569	20
13	-0.114	0.579	21
37-1	1.25	0.6	22
21	-0.077	0.616	23
15-2	0.75	0.74	24
26-2	0.32	0.81	25
16	0.198	0.891	26
33-1	0.62	0.9	27
37-2	0.09	0.91	28
30	0.252	0.945	29
36	0.295	0.988	30
38	0.305	0.998	31
4	0.353	1.046	32
15-3	0.32	1.09	33
35	0.464	1.157	34
9	0.498	1.191	35
5-2	0.05	1.2	36
26-3	1.65	1.29	37
10	0.64	1.333	38
32	0.647	1.34	39
19	0.796	1.489	40

Table A1 - Ordered Booklet Spreadsheet: Reading 2003
(continued)

Item	Rasch Difficulty	Ach @ 2/3 Pass	Page
37-3	1.53	1.51	41
8-2	0.63	1.58	42
22	0.893	1.586	43
31	0.896	1.589	44
26-4	-0.01	1.59	45
23	0.934	1.627	46
25	0.957	1.65	47
15-4	1.13	2.04	48
37-4	0.86	2.08	49
29-2	1.28	2.12	50
33-2	1.56	2.41	51

Translating bookmarks into cut scores is the heart of the procedure. Raters sometimes have difficulty with the concept, particularly with the mathematics of it. We will use extreme caution and work through this concept carefully with both groups. Considerable attention is given to this issue in the presentation titled “The Bookmark Procedure” in Attachment B.

Training the Raters. The standard setting session will consist of a full day of training, test administration, and scoring, followed by two days of identifying, debating, revising, and setting standards. A sample agenda is provided below.

April 27, 2004

- 8:00 a.m. Registration, Materials, Refreshments
- 8:30 Introductions; Collection of Security Forms
- 8:45 Background and Overview
- 10:00 Break
- 10:15 Test Administration
- 12:30 Lunch
- 1:30 p.m. Test Scoring and Discussion
- 3:00 Review of Achievement Levels
- 4:00 Adjourn

April 28, 2004

8:00 a.m. Materials, Refreshments
8:30 Introduction to the Bookmark Procedure
10:00 Break
10:15 Practice Test
11:00 Questions & Answers
Noon Lunch
1:00 p.m. Instructions for Round 1
1:15 Round 1
3:45 Wrap-up
4:00 Adjourn

April 29, 2004

8:00 a.m. Materials, refreshments
8:30 Review of Round 1
9:45 Round 2
Noon Lunch
1:00 p.m. Discussion of Round 2
1:30 Round 3
3:00 Final recommendations
3:30 Closure
4:00 Adjourn

Drs. Bunch and Inman will train raters to complete the following tasks:

1. Study and answer all test items in specially ordered test booklets;
2. Identify the knowledge and skills required to answer each item correctly;
3. Determine why later items are more difficult than earlier items in the difficulty-ordered test booklet;
4. Consider whether most of the students just barely performing at the Basic, Proficient, and Advanced levels of achievement would answer the item correctly;
5. Place a mark for that level at each point where levels change;

In the course of the three days of training and rating activity, Drs. Bunch and Inman will perform the following tasks:

1. Provide guidance and feedback throughout the three rounds of standard setting with regard to procedure only;
2. Collect all materials at the end of each round, analyze data, and prepare reports for the next round of standard setting;
3. Provide impact and other data and facilitate discussions between rounds;
4. Provide feedback at the end of the third round to let all participants know

what the final recommendations are likely to be.

Orientation. Dr. Bunch will lead the Reading group, while Dr. Inman will lead the Mathematics group. By the end of the orientation, participants will understand the purpose of standard setting, their role in it, the meaning of the achievement level descriptors, the contents of the tests, and the specific procedures they will follow in setting standards.

Orientation basically consists of preparing raters to answer the following three questions about each test item:

1. What knowledge or skill is required to earn this point?
2. What makes later items more difficult than earlier items?
3. Think of a large group of students at the cut score for this level. Would about 2/3 of them earn this point?

By providing examples and engaging raters in a discussion of those examples, we will bring all participants to a good working knowledge of their tasks. With specific regard to the third question above, Drs. Bunch and Inman will take great care to explain the importance of the 2/3 rule. As noted previously, we will also make sure that the raters are fully aware of their advisory role (hence the title “rater” rather than “standard setter”) and the fact that actual performance standards will be set by the State Board of Education, based on recommendations of the panelists.

Discussion and feedback. Drs. Bunch and Inman will present information from a variety of sources (item difficulty, impact data, results of rating) but will not offer opinions or otherwise attempt to sway raters. They will, however, use the discussion and feedback sessions to ascertain the level of task comprehension of the raters. Should they detect confusion or lack of clarity, they will provide additional training to make sure everyone stays on task. All raters will complete a Readiness Form prior to each round of rating, acknowledging that they have participated in training and that they understand the task they are about to perform.

Test Administration. All raters will take the standard form of the OGT in either Reading or Mathematics. Drs. Bunch and Inman will administer the tests and provide scoring guides (the multiple-choice key plus scoring guides for the constructed-response items). We note that the scoring of CR items is a very involved process requiring days of training for MI scorers. Raters will not receive the full range of scorer training. Instead, we will present an abbreviated version of scorer training and allow raters to score their own tests. After they have taken the tests and scored their own papers, raters will discuss the overall difficulty and contents of the tests.

The scoring guides deserve some attention at this point because they may well influence raters. As noted above, MI scorers receive extensive guides and days of training before they score student responses to CR items. We will not have several

days to train raters to be fully qualified scorers, nor would it be appropriate to attempt to train them to be scorers. The student responses they will eventually review will have already been scored by qualified scorers. Their own responses (i.e., the ones they will actually “score”) will hardly be representative of high school student responses, for whom the scoring guides and sample papers were developed. Instead, our goal is to give each rater a basic understanding of the requirements of a score point of 1, 2, 3, and so on. During training, we will provide two or three examples of each score point for each CR item (rather than the dozens of sample responses scorers review). Raters will then determine which exemplars are most like their responses in order to score their own tests. Since we will not take high school diplomas away from teachers who score low on the OGT, the main purpose of the score will be to give raters a solid understanding of the overall difficulty of the test.

Practice Test. Our agenda shows a one-hour block of time the morning of the second day for a practice test. The purpose of the practice test will be to give the panelists some experience in setting cut scores before they begin the task with the operational test. The practice test will be a six-item difficulty-ordered booklet with four MC and two CR items. Together, they will span the difficulty of the test; therefore, they will be very distinct from one another. Per previous TAC recommendation, these six items will not come from the actual spring 2004 test, so that there will be no contamination of later ratings of the same items.

Raters will be asked to review the six-item difficulty-ordered practice test booklet and place a bookmark to separate *Basic* from *Proficient*. We have selected this cut point because it is also the cut point for the graduation standard. We will give participants an opportunity to work in small groups to review the six items and identify a point that would separate those just barely qualified to graduate from those not qualified to graduate. The pages of the practice test booklet will contain the same information as those of the regular difficulty-ordered test booklet.

After each rater has placed his or her single bookmark, Drs. Bunch and Inman will ask for a show of hands, tally the locations of the bookmarks, and conduct a brief discussion of their distribution. We will discuss the locations of the bookmarks in terms of their relation to the printed achievement level descriptors and dispersion. We will not share impact data at this point because the goal of the exercise will be to make sure participants fully understand the procedure, not to divide students into categories.

Rating Test Items. At the end of the discussion, we will ask if there are any questions about the task of rating items. We will then distribute the Readiness Form (see Attachment A). Participants will complete the appropriate section of this form and turn it in just before lunch on the second day. We will review the forms during lunch. After lunch, if any participants have indicated that they are not ready to proceed, we will provide further training. If one or two participants continue to

struggle, we will work individually with them after distributing Round 1 materials and instructions to the rest of the participants.

Presenting the difficulty-ordered test booklet. We will use the spring 2004 operational tests as the basis for the difficulty-ordered test booklets. In preparing the difficulty-ordered booklet, MI will include the page number, the original item number, and the Rasch achievement level associated with a 2/3 probability of answering a given MC item correctly or obtaining that score point or better for a CR item. Raters will know at a glance how difficult the item was. They will not know, simply from the information on the page, what the cut score would be or how many students would score above or below the cut score.

Figure A1 shows the layout of a hypothetical difficulty-ordered booklet. The bold numbers at the top right of each page indicate the position in the difficulty-ordered booklet. The numbers at the top left indicate the positions in the original booklet. Some of these numbers have hyphens (e.g., 21-1). These numbers refer to the original item number (21, in this case) and the score point represented on that page (1, in this case). Each such item will appear in the booklet once for each of its score points.

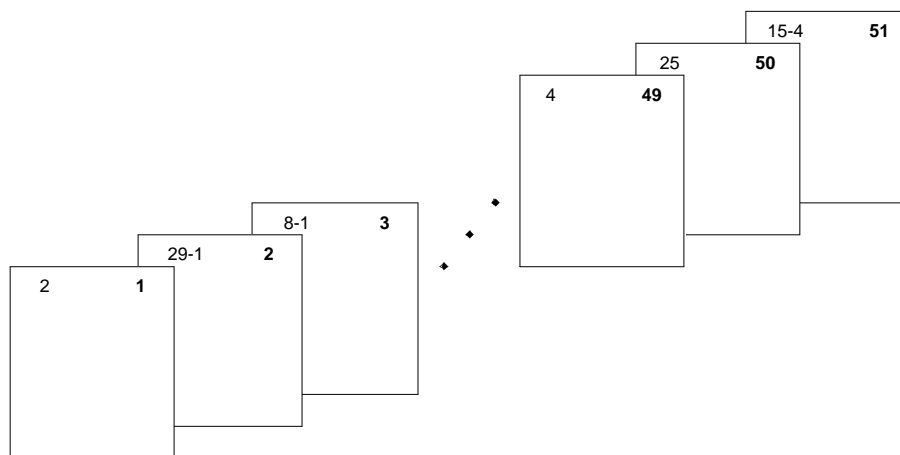


Figure A1. Hypothetical page arrangement of a difficulty-ordered booklet

Figure A2 shows an enlargement of a single item page that includes page number, original item number and score point, page number, and the Rasch achievement level required for a 2/3 chance to answer the item correctly. The key (A) is placed at the bottom of the page in a smaller font to serve as a quick check on the rater's own response to the item without interfering with the rater's estimation of the difficulty of the item. In practice, since items associated with a given reading passage are likely to vary widely in difficulty and therefore be scattered throughout the test booklet, all passages and other common stimulus materials are placed in a companion booklet. The companion booklet is distributed to raters along with the difficulty-ordered test booklet. In practice, raters gain enough familiarity with the passages during the first day of test administration and scoring and the first round

of item rating that they typically refer to the companion booklet very little during the last two rounds of item rating.

1
Item 2
Achievement level required for a 2/3 chance to answer correctly: $-.67$
Which of these best supports the idea that Mary McLeod Bethune is concerned with helping young people find their way in the world?
A. the legacy she leaves in her will B. her desire to return and help Essie C. her zeal for her own place in history D. the way she inspires Essie to believe
Key = A

Figure A2. Sample page from a difficulty-ordered booklet

The page shown in Figure A2 contains all the information a rater would need to make a decision about the item. This item, on page 1 of the difficulty-ordered booklet, was item 2 in the spring 2004 test. The achievement level required for a student to have a 2/3 chance of answering this item correctly is $-.67$ (the Rasch difficulty value plus $.693$, rounded to two decimal places). All the identifying information is at the top of the page so that it will be easily accessible to raters.

The page number will be bold and of a larger size to make it clearly distinguishable from the other numbers. We will ask raters to use the page number as their indicator for a bookmark.

Round 1. Each rater will receive a difficulty-ordered test booklet, passage booklet (for Reading only), and a set of bookmarks. The difficulty-ordered booklet, constructed in accordance with Table A1, will have one item per page, starting with the easiest item in the test booklet. At the top of each page will be printed the original item number (top left), page number (top right), item Rasch difficulty index (top center), and Rasch achievement level associated with 2/3 chance of answering correctly for MC items or 2/3 chance of obtaining that raw score or higher for CR items. Each CR item will be represented once for each of its score points, as noted above. We will print the item and one sample response per score point. Because there will be several different ways to earn each score point, we will select sample responses to cover the full range of possibilities across the various CR items.

For the Reading test, we will prepare a separate passage booklet. In a difficulty-ordered test booklet, items for a given passage will be scattered throughout the booklet. Raters sometimes have some difficulty with this feature of the procedure, and we try to help them in any way we can. We will create a meaningful code for each passage and then repeat that code at the beginning of each associated item. Thus, if a rater needs to refer to a particular passage, that rater would have no difficulty in doing so.

The bookmarks will be printed on one side of a piece of card stock. Each bookmark will be similar to the one shown in Figure A3. In Rounds 1 and 2, raters will enter the page number for each bookmark. At Round 3, raters will be familiar with the relationship between page number and cut score. We will ask them to enter page number and associated cut score, as well as the impact data in order to make sure each rater is fully aware of his or her recommendation. In introducing the bookmark, we will emphasize the fact that the ODE has designated *Proficient* as the graduation standard. Therefore, we will direct raters to focus on this standard first and then turn their attention to *Basic*, *Accelerated*, and *Advanced*.

During Round 1, raters will work in small groups of 3-5 individuals. While they will discuss the item contents among themselves, each rater will complete his or her own bookmark card. As they complete Round 1, raters will review their forms to make sure they are complete, return all materials to Dr. Bunch or Dr. Inman, and be dismissed for the day.

OGT Standard Setting
Mathematics

Rater Number _____

Bookmarks (Enter Page Number for Rounds 1 and 2.)

Round	Basic	Proficient/ Graduation	Accelerated	Advanced
1				
2				

Round 3

	Basic	Proficient/ Graduation	Accelerated	Advanced
Page Number				
Cut Score				
% At or Above				

Notes

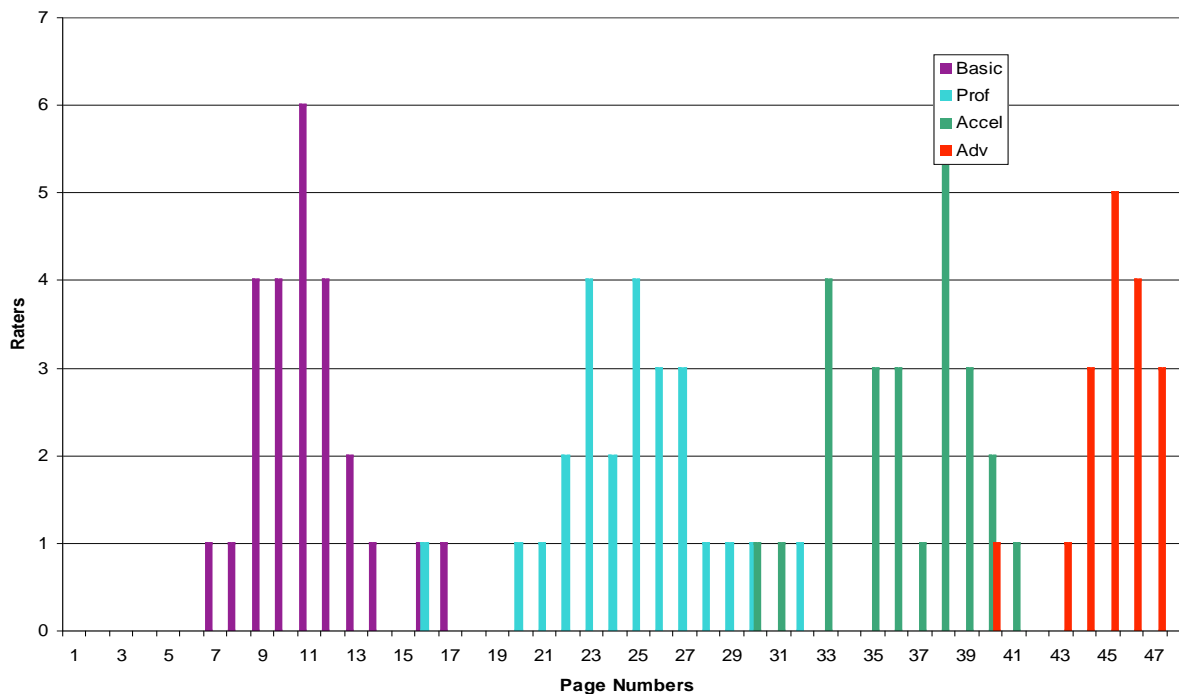
Figure A3. Bookmark Reporting Form

Data analysis and presentation. At the end of each round, MI staff will collect the bookmark cards and enter the values into a spreadsheet similar to the one shown in Table A2 and a chart similar to the one shown in Figure A4. After tallying the results, we will return the cards to the raters, along with the results. Table A2 allows raters to see where their bookmarks fall, relative to those of other raters. It also gives them a sense of where the group average lies, as well as how far their own bookmarks fall from the group average. For our purposes here, Figure A4 provides a more graphic summary of the bookmarks in the region of the Proficient/Accelerated division, allowing each rater to see his or her marks, relative to those of other raters.

Table A2 Sample Output for One Round of Standard Setting

Results of Round 1 of Rating								
Rater	Basic		Proficient/Graduation		Accelerated		Advanced	
	Page	Ach.	Page	Ach.	Page	Ach.	Page	Ach.
1	11	0.273	27	0.900	36	1.200	48	2.040
2	9	0.193	23	0.616	31	0.998	40	1.489
3	13	0.420	25	0.810	30	0.988	49	2.080
4	9	0.193	25	0.810	40	1.489	47	1.650
5	12	0.286	26	0.891	38	1.333	46	1.627
6	7	-0.176	20	0.569	33	1.090	44	1.589
7	16	0.493	23	0.616	41	1.510	45	1.590
8	8	0.082	27	0.900	40	1.489	45	1.590
9	10	0.270	27	0.900	38	1.333	45	1.590
10	12	0.286	24	0.740	38	1.333	48	2.040
11	11	0.272	21	0.579	38	1.333	46	1.627
12	11	0.272	22	0.600	36	1.200	44	1.589
13	10	0.270	16	0.493	35	1.191	45	1.590
14	17	0.540	23	0.616	33	1.090	46	1.627
15	14	0.440	26	0.891	39	1.340	48	2.040
16	12	0.286	32	1.046	39	1.340	49	2.080
17	10	0.270	29	0.945	38	1.333	47	1.650
18	9	0.193	26	0.891	33	1.090	43	1.586
19	11	0.272	22	0.600	39	1.340	44	1.589
20	11	0.272	25	0.810	36	1.200	45	1.590
21	12	0.286	24	0.740	37	1.290	50	2.120
22	13	0.420	28	0.910	35	1.191	48	2.040
23	11	0.272	23	0.616	35	1.191	48	2.040
24	10	0.270	25	0.810	33	1.090	47	1.650
25	9	0.193	30	0.988	38	1.333	46	1.627
Mean		0.274		0.771		1.253		1.749
SD		0.138		0.156		0.145		0.220
M-1SD		0.136		0.616		1.108		1.529
M+1SD		0.412		0.927		1.397		1.970
Mean Cut		25.0		32.5		38.5		42.5
Low Cut		20.0		28.0		35.0		41.0
Cut-1SD		23.5		30.0		36.0		41.0

Cut+1SD		27.0		34.0		39.0		42.5
---------	--	------	--	------	--	------	--	------



High Cut		29.0		35.0		41.0		46.0
----------	--	------	--	------	--	------	--	------

Figure A4. Data presentation for one round of standard setting showing the Proficient/Accelerated division

Table A2 provides a summary of bookmark placements, as well as the resulting cut scores. Note that the mean cut score is shown (along with its standard deviation). Also shown are the low cut, high cut, and cuts one standard deviation above and below the mean. Individual cut scores are not shown. The raters already know those values, having noted them on the pages where they placed their bookmarks. We use Table 2 to draw attention to student achievement levels and the associated mean cut scores. It is also noteworthy that the mean cut score based on the mean Rasch achievement level is not necessarily equal to the arithmetic mean of the individual cut scores. Rather than have to explain why they are not (or even worse, respond to someone’s conclusion that we miscalculated the mean cut score), we choose to report bookmarks and associated abilities at the individual rater level and save cut score for the summary.

Figure A4 offers a different perspective. Here, only the bookmarked page numbers are shown. Raters get a graphic view of how their bookmarks compare to the bookmarks of the other raters. Figure 4 also shows where there are gaps; i.e., page ranges no one chose as bookmarks for any cut score. In subsequent rounds, these

pages will typically not enter the discussion. This figure also shows where there are overlaps. For example, one rater chose page 32 as the bookmark for *Proficient*, while another rater chose 30 and another 31 for *Accelerated*. In effect, one rater would set the cutoff for *Proficient* higher than at least one rater would set the cutoff for *Accelerated*. Visualizations such as Figure A4 are excellent conversation starters in Rounds 2 and 3.

Impact data. MI will show impact data to raters between Rounds 1 and 2. The purpose of the data will be to allow raters to see how many students would be classified at each achievement level if the mean cut scores from Round 1 were implemented. Given the timing of the standard-setting activity, we will present impact data for a representative sample of students ($N \approx 40,000 \text{ } \alpha \text{ } 50,000$) who take the test as well as for subgroups based on gender and ethnicity. We plan to present these data in both tabular and graphic form, as shown in Table A3 and Figures A5 and A6.

Actual data from the spring 2003 administration of the OGT Reading test are shown in Table A3. For 2004, we will use spring 2004 test results. In this table, each score point is treated as a potential cut score. The four cut scores established in Table A2 (rounded to whole numbers) are highlighted in Table A3. Numbers of students obtaining each score point are shown, along with cumulative totals through each score point. Percentages of students obtaining each score point and cumulative percentages are also given. The Inverse column shows the percentage of students scoring at or above each score point. This column is added to answer the first question that is usually asked, “How many students passed?” Using this column, raters can quickly see what percentage of students “pass” at any given cut score. This column is depicted graphically in Figure 6.

Table A3 - Impact Data: Reading (Round 1)

Ohio Graduation Test							
Impact Data: Reading							
Cut Score	Freq.	Cum. Freq.	Percent	Cum. %	% At or Above	Status	% In Group
0	16	16	0.01	0.01	100.00	Limited	23.53
1	0	16	0.00	0.01	99.99	Limited	
2	8	24	0.01	0.02	99.99	Limited	
3	33	57	0.02	0.04	99.98	Limited	
4	53	110	0.04	0.08	99.96	Limited	
5	81	191	0.06	0.13	99.92	Limited	
6	101	292	0.07	0.20	99.87	Limited	
7	277	569	0.19	0.39	99.80	Limited	
8	308	877	0.21	0.61	99.61	Limited	
9	413	1290	0.29	0.89	99.39	Limited	
10	597	1887	0.41	1.30	99.11	Limited	
11	892	2779	0.62	1.92	98.70	Limited	
12	1032	3811	0.71	2.63	98.08	Limited	
13	1176	4987	0.81	3.45	97.37	Limited	
14	1301	6288	0.90	4.34	96.55	Limited	
15	1500	7788	1.04	5.38	95.66	Limited	
16	1701	9489	1.18	6.56	94.62	Limited	
17	1904	11393	1.32	7.87	93.44	Limited	
18	2184	13577	1.51	9.38	92.13	Limited	
19	2544	16121	1.76	11.14	90.62	Limited	
20	2923	19044	2.02	13.16	88.86	Limited	
21	3201	22245	2.21	15.37	86.84	Limited	
22	3322	25567	2.30	17.66	84.63	Limited	
23	3875	29442	2.68	20.34	82.34	Limited	
24	4612	34054	3.19	23.53	79.66	Limited	

Table A3 - Impact Data: Reading (Round 1)
(Continued)

Ohio Graduation Test							
Impact Data: Reading							
Cut Score	Freq.	Cum. Freq.	Percent	Cum. %	% At or Above	Status	% In Group
25	5219	39273	3.61	27.13	76.47	Basic	32.91
26	5887	45160	4.07	31.20	72.87	Basic	
27	5833	50993	4.03	35.23	68.80	Basic	
28	5904	56897	4.08	39.31	64.77	Basic	
29	6086	62983	4.20	43.52	60.69	Basic	
30	6308	69291	4.36	47.87	56.48	Basic	
31	6245	75536	4.31	52.19	52.13	Basic	
32	6160	81696	4.26	56.44	47.81	Basic	
33	6097	87793	4.21	60.66	43.56	Proficient	23.92
34	6027	93820	4.16	64.82	39.34	Proficient	
35	5833	99653	4.03	68.85	35.18	Proficient	
36	5714	105367	3.95	72.80	31.15	Proficient	
37	5522	110889	3.82	76.61	27.20	Proficient	
38	5426	116315	3.75	80.36	23.39	Proficient	
39	5309	121624	3.67	84.03	19.64	Accelerated	12.64
40	5087	126711	3.51	87.55	15.97	Accelerated	
41	4114	130825	2.84	90.39	12.45	Accelerated	
42	3782	134607	2.61	93.00	9.61	Accelerated	
43	3102	137709	2.14	95.14	7.00	Advanced	7.00
44	1633	139342	1.13	96.27	4.86	Advanced	
45	1504	140846	1.04	97.31	3.73	Advanced	
46	1416	142262	0.98	98.29	2.69	Advanced	
47	1163	143425	0.80	99.09	1.71	Advanced	
48	869	144294	0.60	99.69	0.91	Advanced	
49	302	144596	0.21	99.90	0.31	Advanced	
50	125	144721	0.09	99.99	0.10	Advanced	
51	16	144737	0.01	100.00	0.01	Advanced	

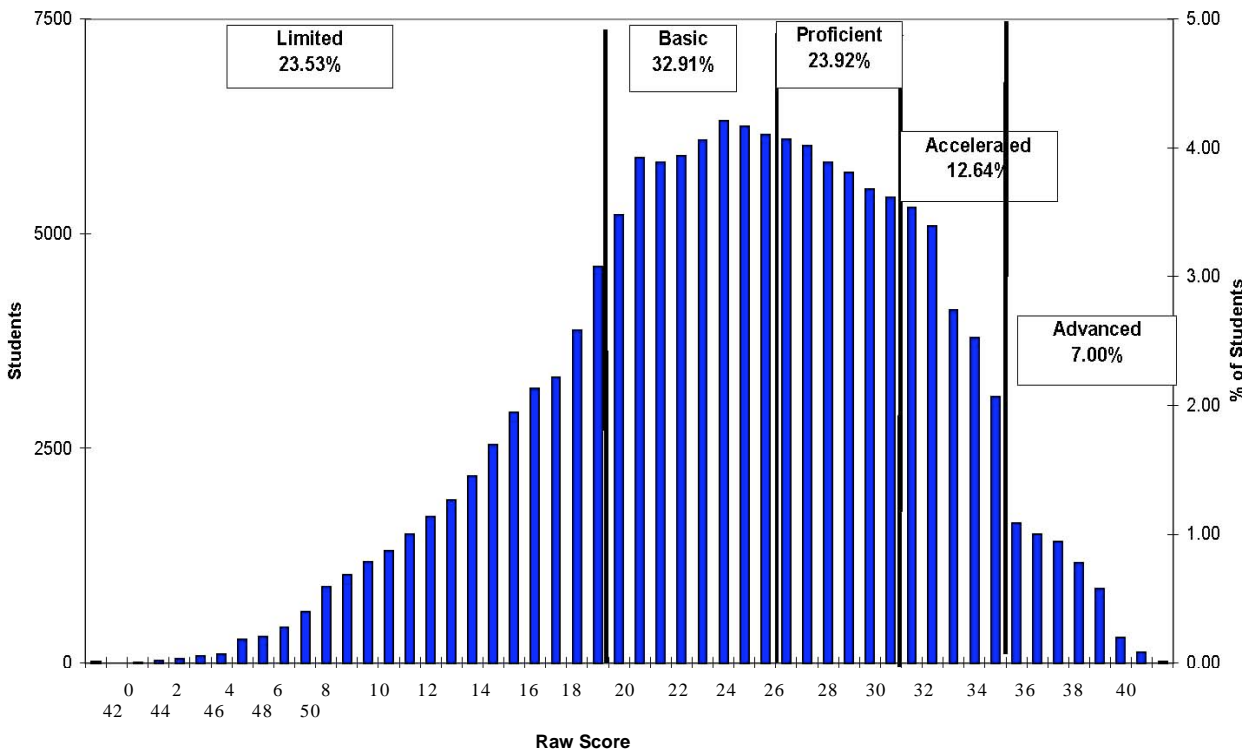


Figure A5. Impact data showing numbers of students at each achievement level

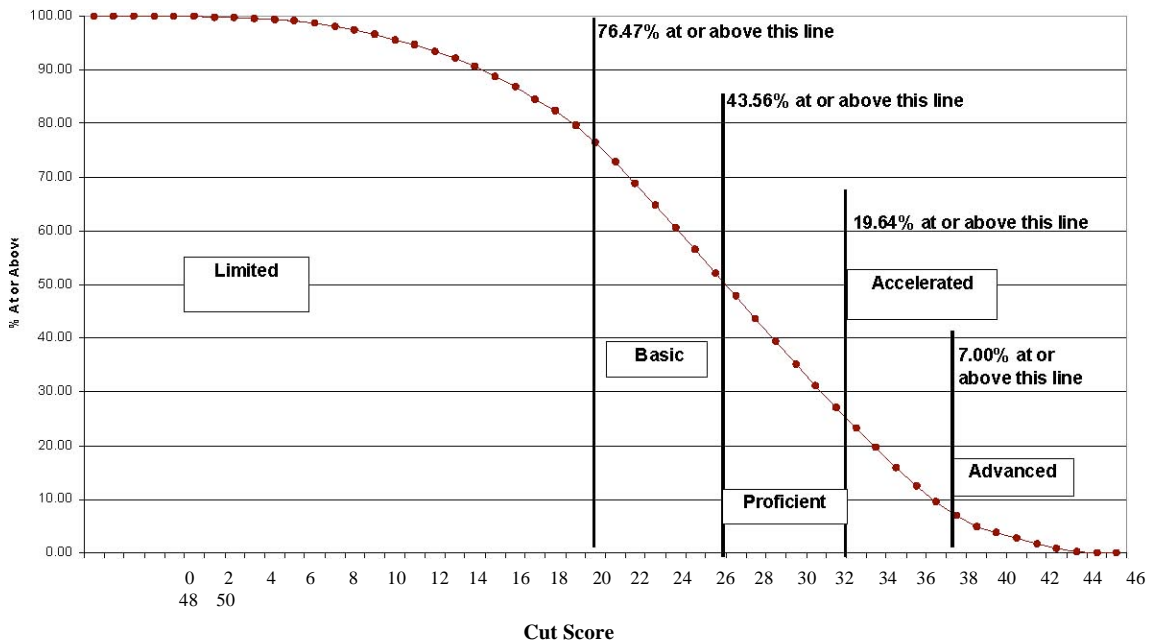


Figure A6. Impact data showing percent of students scoring at or above each cut score

All the data for Table A3 and Figures A5 and A6 will be in hand prior to the

standard setting session. Only the last two columns of Table A3 will need to be adjusted after each round to superimpose the mean ratings on the existing score distributions. Similarly, Figures 5 and 6 will be predrawn except for the cut lines and reported percentages (percent in group and percent at or above this line). Thus, these reports to the raters will be both helpful and easy to produce quickly.

Round 2. On the morning of Day 3, raters will receive their difficulty-ordered test booklets and other materials from Round 1 plus the data from Round 1 and the impact data. Drs. Bunch and Inman will lead a discussion of the Round 1 ratings and impact. Table A3 and Figures A4, A5, and A6 on the preceding pages show the types of information we will share with raters and the manner in which we will present it.

Discussion will focus on range of cut scores, areas of particular disagreement, and concerns about placement of individual items. It is sometimes helpful for raters to work through the difference between their perceived difficulty of a particular item and the placement of that item relative to others in the test. Once raters have discussed the results of Round 1 as a total group, we will reassign them to small groups of 3 to 5 members each, and begin Round 2. The task for Round 2 will be identical to that of Round 1. The primary difference will be the amount of information available to each rater. At the end of the round, Drs. Bunch and Inman will collect all materials and dismiss for lunch.

During lunch, Drs. Bunch and Inman will once again tally the cut scores and prepare reports similar to those shown in Tables A2 and A3 and Figures A4, A5 and A6. They will present these to participants at the beginning of Round 3.

Round 3. As they return from lunch, raters will receive all their Round 2 materials plus a summary of Round 2. We will use a special form of Table A1 which includes actual cut scores associated with the values of Rasch achievement level (see Table A4). Once again, raters will be able to examine the impact data as part of their discussion. Drs. Inman and Bunch will lead a discussion of the impact data and other topics of concern from Round 2. At the end of this discussion, raters will have a final opportunity to evaluate all their previous ratings and all information at hand and simply enter four bookmarks and the associated cut scores. The cut scores will come from a table similar to Table A4. We will also ask raters to enter the % At or Above each cut score they enter. We have added this value as a check on the accuracy of the cut score entries. Drs. Bunch and Inman will check each completed bookmark for accuracy, tally these final ratings and calculate the mean recommended cut score for each achievement level. At the discretion of the ODE, we can report these means to the raters or not.

Table A4
Round 3 Ordered Booklet Spreadsheet

OGT Reading Ordered Booklet: Spring 2003					
Item	Rasch Difficulty	Ach. @ 2/3 Pass	Page	Associated Cut Score	% At or Above
6	-2.305	-1.612	1	8	99.61
2	-1.986	-1.293	2	10	99.11
3	-1.95	-1.257	3	10	99.11
12	-1.304	-0.611	4	15	95.66
14	-1.027	-0.334	5	18	92.13
11	-0.936	-0.243	6	19	90.62
1	-0.869	-0.176	7	20	88.86
28	-0.611	0.082	8	23	82.34
24	-0.5	0.193	9	24	79.66
15-1	0.48	0.27	10	25	76.47
18	-0.421	0.272	11	25	76.47
17	-0.407	0.286	12	25	76.47
26-1	0.79	0.42	13	27	68.80
5-1	0.65	0.44	14	27	68.80
8-1	0.35	0.44	15	27	68.80
7	-0.2	0.493	16	28	64.77
29-1	0.24	0.54	17	29	60.69
27	-0.142	0.551	18	29	60.69
34	-0.135	0.558	19	29	60.69
20	-0.124	0.569	20	29	60.69
13	-0.114	0.579	21	29	60.69
37-1	1.25	0.6	22	30	56.48
21	-0.077	0.616	23	30	56.48
15-2	0.75	0.74	24	32	47.81
26-2	0.32	0.81	25	33	43.56
16	0.198	0.891	26	34	39.34
33-1	0.62	0.9	27	34	39.34
37-2	0.09	0.91	28	34	39.34
30	0.252	0.945	29	35	35.18
36	0.295	0.988	30	35	35.18
38	0.305	0.998	31	35	35.18
4	0.353	1.046	32	36	31.15
15-3	0.32	1.09	33	36	31.15
35	0.464	1.157	34	37	27.20
9	0.498	1.191	35	38	23.39
5-2	0.05	1.2	36	38	23.39
26-3	1.65	1.29	37	39	19.64

10	0.64	1.333	38	39	19.64
----	------	-------	----	----	-------

Table A4
Round 3 Ordered Booklet Spreadsheet
 (continued)

OGT Reading Ordered Booklet: Spring 2003					
Item	Rasch Difficulty	Ach. @ 2/3 Pass	Page	Associated Cut Score	% At or Above
32	0.647	1.34	39	39	19.64
19	0.796	1.489	40	41	12.45
37-3	1.53	1.51	41	41	12.45
8-2	0.63	1.58	42	42	9.61
22	0.893	1.586	43	42	9.61
31	0.896	1.589	44	42	9.61
26-4	-0.01	1.59	45	42	9.61
23	0.934	1.627	46	42	9.61
25	0.957	1.65	47	42	9.61
15-4	1.13	2.04	48	45	3.73
37-4	0.86	2.08	49	45	3.73
29-2	1.28	2.12	50	46	2.69
33-2	1.56	2.41	51	47	1.71

Table A4 is a modified version of Table A1 with two extra columns. The final two columns of Table A4 translate student achievement levels into cut scores (Associated Cut Score) and then percentages of students scoring at or above that cut score. The Associated Cut Score is from the Rasch analysis output from BIGSTEPS/WINSTEPS table that associates a Rasch ability level with each raw score point on the test. The final column (% At or Above) is taken directly from Table A3 and is added here so that all necessary information is in one place.

The rationale for presenting Table A4 and asking panelists for raw cut scores in addition to bookmarks is simple. By the afternoon of the third day of review of the tests, panelists will have processed a considerable amount of information about item difficulty, student achievement, associated cut scores, and percentages of students scoring at or above those cut scores. We will have presented some of the information in more than one format. By Round 3, we believe panelists will be ready to make a direct translation from their bookmarks (reasonably fixed by this time) to raw cut scores. Table A4 will allow them to make and check those translations. The added benefit of having panelists give raw cut scores on the final round will be that they will have a clearer sense of the impact of their actions than if they simply give page numbers. As noted above, we will ask panelists not only to enter page numbers and raw cut scores but percentages of students scoring at or above those cut scores as well. This last measure is to allow us to make sure panelists have entered the numbers they meant to enter, and it is a final reminder to the panelists of the impact of their individual decisions.

Follow-Up and Reporting. MI staff will summarize the processes and outcomes of the three-day session and present the plans and results to ODE, the TAC, and the TSC and FSRC. It is likely that one or more of the groups will have comments about the results and may even want to alter the proposed cut scores. To preserve the integrity of the overall process, we propose to append the recommendations of these groups to the original set of recommendations and forward all to the State Board of Education.

Drs. Bunch and Inman will work with ODE staff to prepare a report for the Ohio Board of Education, complete with recommendations regarding cut scores. The report will provide a complete description of the process as well as the cut scores recommended by the raters during the three-day activity as well as any additional recommendations from TAC, TSC, or FSRC. The report will also contain an executive summary for nontechnical audiences. Either Dr. Bunch or Dr. Inman will be available for a face-to-face meeting with the Board, if ODE so chooses.

References

- Lewis, D. M., Mitzel, H. C. & Green, D. R. (1996). Standard setting: A bookmark approach. In D. R. Green (Chair), IRT-Based Standard Setting Procedures Utilizing Behavioral Anchoring. Symposium conducted at the Council of Chief State School Officers National Conference on Large-Scale Assessment, Phoenix, AZ.
- Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The bookmark procedure: Psychological perspectives. In G. Cizek (Ed.), Setting Performance Standards: Concepts, Methods , and Perspectives. Mahwah, NJ: Lawrence Erlbaum & Associates.
- Wright, B. D. & Stone, M. H. (1979). Best Test Design. Chicago: MESA Press.
- Wright, B. D. & Masters, G. N. (1982). Rating Scale Analysis. Chicago: Mesa Press.

Appendix B: Sample Standard Setting Materials

Follow-Up Letters

Outlines for Rounds 1 – 3

Readiness Form

Workshop Evaluation

Preparation Schedule

Materials List

Achievement Level Descriptors (Draft)

Follow up letter

Dear

Thank you for agreeing to serve on the standard setting committee for Ohio Graduation Test (OGT). Materials you will need are enclosed. These include directions to XXX and the meeting room location, an agenda, and a brief explanation of the procedure we will follow. Please bring these with you when you come.

Responsibilities of raters include the following:

1. Participate in training and practice exercises.
2. Follow instructions provided by the workshop leader.
3. Contribute to group discussions about standards and achievement levels
4. Return all secure materials when asked to do so.
5. Abide by the conditions of the security agreement.

Prior to your reviewing the test booklets, we will ask you to sign a security agreement. A copy is enclosed. By signing this agreement, you agree not to divulge the contents of any test or test item to anyone without written permission from the Ohio Department of Education (ODE). Please bring this agreement with you as well.

I look forward to seeing you in Columbus.

Sincerely,

William R. Batchelor, Ph.D.
Senior Project Director

C: Mr. Tom Bulgrin, ODE

Ohio Graduation Tests
Standard Setting Meeting
April 27 - 29, 2004

Agenda

April 27

- 8:00 a.m. Registration, Materials, Refreshments
- 8:30 Introductions
- 8:45 Background and Overview
- 10:00 Break
- 10:15 Test Administration
- 12:00 Lunch
- 1:00 p.m. Test Scoring and Discussion
- 3:00 Review of Performance Level Descriptors
- 4:00 Adjourn

April 28

- 8:00 a.m. Materials, Refreshments
- 8:30 Introduction to the Bookmark Procedure
- 10:30 Practice Test
- 11:30 Questions & Answers
- Noon Lunch
- 1:00 p.m. Instructions for Round 1
- 1:15 Round 1
- 3:45 Wrap-up
- 4:00 Adjourn

April 29

- 8:00 Materials, refreshments
- 8:30 Review of Round 1
- 9:45 Round 2
- Noon Lunch
- 1:00 p.m. Discussion of Round 2
- 1:30 Round 3
- 3:00 Final recommendations
- 3:30 Closure
- 4:00 Adjourn

During these three days, you will have an opportunity to examine, item-by-item, the OGT. You will review test booklets in which the test items have been rearranged in order of difficulty, with the easiest item first and the most difficult item last. After a group discussion of what it means for a high school student to be at Limited, Basic, Proficient, Accelerated, or Advanced levels of achievement, you will examine the items in your test booklet to place three bookmarks. The first point will separate students performing at a Limited level from those at the Basic level of achievement. The second point will separate students at the Basic level from those at the Proficient level. A third point will separate students at the Proficient level from those at the Accelerated level, and the fourth will separate students at the Accelerated level from those at the Advanced level of achievement.

Having marked your test, you will have an opportunity to discuss your cutoff points with others. We will also present test item information and give you two more opportunities to consider the placement of the bookmarks, with the benefit of further discussion with other committee members and information about how students actually performed on the test.

Outline for Round 1

Raters will work in groups of 3 to 5 people.

1. Distribute the following to participants:
 - a. Test
 - b. Prompts, passages, and common material
 - c. Test spreadsheet
 - d. Bookmark
2. Direct participants to enter their rater # on each of the 4 things listed above
3. Divide into groups of 5 people, any method is OK.
4. Give directions:
 - a. In small groups, examine a small cluster (5 or 6 items) individually and then discuss the content necessary to answer item correctly. Make notes about content/skills measured in the test booklet. Try to answer item without the key and without the other group members before making that determination. Remember that these items have been reviewed several times by Ohio educators whose ideas about what the item actually measures may differ from yours. If you have a comment about the content validity of an item, mark it in the booklet. We will pass it along to ODE. There is no need to introduce this information into the group discussion.
 - b. Discuss what makes items in this cluster more difficult than those in previous clusters.
 - c. Repeat steps 4a and 4b until you have reached the end of the booklet.
 - d. Using the notes you have made during your review and discussions, go back through the entire test on your own and try to locate the group of items where the cut score between Basic and Proficient might lie. For each item in this vicinity ask yourself, "If a large group of students who were just barely Proficient were to attempt this item, would about 2/3 of them get it right (or get this score point)?" If the answer is yes, go on to the next item. If the answer is no, go back to the last item for which the answer is yes. Mark this item's page number on your bookmark as the dividing line between the Basic and Proficient levels.
 - e. Do the same thing for Proficient/Accelerated.
 - f. Do the same thing for Accelerated/Advanced.
 - g. Do the same thing for Limited/Basic.
 - h. Make a final decision as to whether to move your bookmarks or leave them where they are.
 - i. Once you are satisfied that you have entered three bookmarks you could defend, place your bookmark on top of your test booklet and return your bookmark, test booklet, test spreadsheet, and prompts, passages, and common materials to your group leader. Leave all your other materials (packets and handouts) on your table. When all your materials are accounted for, you may leave.
5. Make sure all materials are accounted for. Acknowledge receipt of materials and

dismiss participants individually or as a group. Check each bookmark to make sure it has a rater number and three valid entries. Give these to the administrative assistant.

Outline for Round 2

Raters will work in groups of 5; assignments will be made on the basis of Round 1 results so that each group will have a range of perspectives.

1. Distribute the following to participants:
 - a. Day 1 packets with rating cards (bookmarks)
 - b. Round 1 graphics
 - c. Impact data
2. Discuss Round 1
 - a. Leader explains graphic, noting overlaps, spread, etc.
 - b. Start discussion with support for middle scores, NOT extremes
 - c. Discuss impact data
 - d. Work out from mean cut scores until all who want to have spoken
3. Introduce Round 2
 - a. Eliminate from consideration items below the lowest Basic cut, above the highest Advanced cut, or in-between scores that were not selected by any rater.
 - b. Repeat Round 1 process of working in small groups and focus on what makes it more difficult to achieve later points than earlier points.
4. Make sure all vouchers are turned in to Bill by noon. If your group finishes Round 2 before noon, secure all materials and give them a long break—resume work after noon lunch break.
5. Collect Round 2 bookmarks and deliver to Mike (Reading) or Elliot (Math).
6. Secure the room for lunch.
7. Pick up checks from Bill before 1 p.m.

Outline for Round 3

Continue to work in groups of 5; these may be the same as in Round 2 or not, depending on the panelists.

1. Distribute Round 2 graphic and impact data.
2. Discuss Round 2 results, convergence, and change in mean (if any).
3. Begin discussion with middle scores and work out. Stop when it is apparent that everyone who wants to talk has done so.
4. Begin Round 3 rating, focusing only on scores still under consideration.
5. Collect all materials and verify completeness. Thank raters for participating, and give them their checks.

Ohio Graduation Tests

Readiness Form

Rater Number _____

Practice Test: I have completed the practice test, and I understand what I need to do to complete Round 1.

(Circle one): **Yes** **No**

Round 1: I have discussed the results of Round 1, including my ratings, the ratings of others, and the impact data, and I understand what I need to do to complete Round 2

(Circle one): **No** **Yes**

Round 2: I have discussed the results of Round 2, including my ratings, the ratings of others, and the impact data, and I understand what I need to do to complete Round 3

(Circle one): **No** **Yes**

Round 3: I have completed my ratings, and I believe that the cut scores I have identified fairly represent minimal performances of students at the Basic, Proficient, and Advanced levels

(Circle one): **No** **Yes**

Everyone was encouraged to share his or her ratings and hear those of other raters.

(Circle one): **No** **Yes**

The cut scores we recommended accurately reflect the basic, proficient, and advanced achievement levels.

(Circle one): **No** **Yes**

The process was fair and unbiased.

(Circle one): **No** **Yes**

Workshop evaluation sheet

**Ohio Graduation Tests
Standard Setting Workshop Evaluation**

Please respond to the statements below by indicating your agreement or disagreement. Check one box for each statement to indicate whether you strongly agree, agree, disagree, or strongly disagree with each statement. Room for comments has been provided at the bottom of the form.

Subject:

(Circle One):

Reading

Mathematics

	Statement	Strongly Agree	Agree	Disagree	Strongly Disagree
1	The workshop leaders clearly explained the purpose of the meeting.				
2	The workshop leaders clearly explained my task.				
3	The examples and exercises helped me understand how to perform my task.				
4	The large and small group discussions helped me understand the process.				
5	I was able to follow the instructions and complete the rating sheets accurately.				
6	The discussions after the first round of rating were helpful to me.				
7	The discussions after the second round of rating were helpful to me				
8	The information showing the distribution of student scores was helpful to me.				
9	The facilities and food service helped to create a good working environment.				

Comments

Signature

Date

Schedule

Ohio Graduation Test			
Spring 2004 Standard Setting Schedule			
Task	Responsibility	Start Date	End Date
Send overall plan to ODE for TAC review	MI	11/10/2003	11/10/03
Present plan to TAC	MI	12/04/2003	12/04/03
Revise plan in accordance with TAC input	MI	12/05/2003	12/12/03
Review revised plan	TAC	12/15/2003	12/19/2003
Revise plan in accordance with TAC input	MI	12/22/2003	2/13/2004
Draft achievement level descriptors	MI/ODE	1/05/2004	2/13/2004
Revise achievement level descriptors	MI/ODE	2/16/2004	2/27/2004
Send revised plan to ODE	MI	2/16/2004	2/16/2004
Review and approve revised plan	ODE	2/17/2004	2/27/2004
Present final plan to TAC	MI	3/04/2004	3/05/2004
Present plan to TSC/FSRC	MI	3/08/2004	3/12/2004
Revise plans based on committee and TSC/FSRC input	MI	3/15/2004	3/26/2004
Send final Reading plan to ODE for review	MI	3/29/2004	4/07/2004
Send final Math plan to ODE for review	MI	3/29/2004	4/07/2004
Create Excel programs to analyze Round 1 and 2 data	MI	3/31/2004	4/07/2004
Create Powerpoint presentations	MI	3/31/2004	4/07/2004
Create scoring guides	MI	3/31/2004	4/07/2004
Submit final comments on plans	ODE	4/03/2004	4/07/2004
Complete hotel arrangements	MI	4/07/2004	4/08/2004
Approve plans	MI	4/08/2004	4/14/2004
Send rooming list to hotel	MI	4/14/2004	4/14/2004
Practice applying scoring guides (with staff)	MI	4/14/2004	4/16/2004
Rehearse presentations	MI	4/15/2004	4/16/2004
Print evaluation forms	MI	4/19/2004	4/19/2004
Revise materials as necessary	MI	4/19/2004	4/21/2004
Make copies of revised materials	MI	4/21/2004	4/22/2004
Construct Math booklet (difficulty ordered)	MI	4/22/2004	4/23/2004
Construct Reading test booklet (difficulty ordered)	MI	4/22/2004	4/23/2004
Print all rating forms	MI	4/22/2004	4/23/2004
Print and copy intro handouts	MI	4/22/2004	4/23/2004
Purchase all supplies (see Supplies)	MI	4/23/2004	4/23/2004
Copy all scoring guides	MI	4/23/2004	4/23/2004
Route copies of final plans to all presenters	MI	4/23/2004	4/23/2004
Assemble rater packets (see Rater	MI	4/23/2004	4/23/2004

Contents)			
Ship materials	MI	4/23/2004	4/23/2004
Conduct final in-house dry run	MI	4/23/2004	4/23/2004
Pack tests, answer documents, administration manuals, math reference sheet, extra specs	MI	4/23/2004	4/23/2004
Get LCD projector from IT	MI	4/26/2004	4/26/2004
Leave for OH	MI	4/26/2004	4/26/2004
Check meeting rooms	MI	4/26/2004	4/26/2004
Check on problems with panelists checking into hotel	MI	4/26/2004	4/26/2004
Conduct Standard Setting	MI/ODE	4/27/2004	4/29/2004
Present results to TSC	MI	5/04/2004	5/04/2004
Present results to FSRC	MI	5/04/2004	5/04/2004
Incorporate recommendations from TAC, TSC, and FSRC	MI	5/05/2004	5/05/2004
Present results to Board	MI/ODE	5/06/2004	5/07/2004
Prepare final report	MI	5/10/2004	5/21/2004
Review final report	ODE	5/24/2004	5/28/2004
Submit final report	MI	6/1/2004	6/11/2004
Approve final report	ODE	6/14/2004	6/25/2004

Materials List

Rater Packet Contents

Agenda
 Notetaking Guide: Overview
 Notetaking Guide: The Bookmark Procedure
 Security Agreement Form
 Readiness Form
 Expense Form
 Bookmark Evaluation Form

Additional Rater Materials

Check	Reading	Check	Mathematics
	Test Booklet		Test Booklet
	Answer Document		Answer Document
	Scoring Key		Scoring Key
	Scoring Guide		Scoring Guide
	Achievement Level Descriptors (4)		Achievement Level Descriptors (4)
	Practice Test		Practice Test
	Difficulty-Ordered Test Booklet		Difficulty-Ordered Test Booklet
	Passage/Stimulus Booklet		
	Round 1 Summary Table		Round 1 Summary Table
	Round 1 Summary Graph		Round 1 Summary Graph
	Round 1 Impact Data Table		Round 1 Impact Data Table
	Round 1 Impact Data Graphs (2)		Round 1 Impact Data Graphs (2)
	Round 2 Summary Table		Round 2 Summary Table
	Round 2 Summary Graph		Round 2 Summary Graph
	Round 2 Impact Data Table		Round 2 Impact Data Table
	Round 2 Impact Data Graphs (2)		Round 2 Impact Data Graphs (2)
	Round 3 Summary Table		Round 3 Summary Table
	Round 3 Summary Graph		Round 3 Summary Graph
	Round 3 Impact Data Table		Round 3 Impact Data Table
	Round 3 Impact Data Graphs (2)		Round 3 Impact Data Graphs (2)

Ohio Graduation Test Achievement Level Descriptors

DRAFT 2/13/03 – To be revised in accordance with HB3

Reading

Basic	Tenth grade students achieving at the Basic level should be able to demonstrate an overall understanding and make some interpretations of informational and literary text. When reading text appropriate to tenth grade, they should be able to do the following: identify and relate aspects of the text to its overall meaning, extend the ideas in the text by making simple inferences, recognize interpretations, draw conclusions, and make connections among and relate ideas in the text to their personal experiences. They should also be able to identify the basic elements of an author's style, as well as how an author appeals to an audience and develops an argument or point of view.
Proficient	Tenth grade students achieving at the Proficient level should be able to show an overall literal and inferential understanding of informational and literary text. When reading text appropriate to tenth grade, they should be able to do the following: extend the ideas in the text by making more sophisticated inferences, conclusions, and connections to their own personal experiences and other readings. Their connections or extrapolations from the text should be clear and plausible, even when implicit. These students should be able to explain the author's use of literary devices, as well as how an author appeals to an audience and develops an argument or point of view.
Advanced	Tenth grade students achieving at the Advanced level should be able to describe, explain, and analyze more abstract themes and ideas in informational and literary text. When reading text appropriate to tenth grade, they should be able to do the following: analyze both the meaning and the form or features of the text, and support their analyses with specific, convincing examples from the text. They should be able to extend the information from the text by relating it in a thoughtful way to their experiences, to other texts, and to the world. Their responses, as well as their level of perception and of interaction with the text should be thorough, mature, and extensive.

Ohio Graduation Test Achievement Level Descriptors

DRAFT 2/13/03 – To be revised in accordance with HB3

Mathematics

Basic	Tenth grade students performing at the Basic level should be able to use estimation to verify solutions and determine the reasonableness of results as applied to real-world problems. Tenth graders performing at the Basic level should recognize relationships presented in verbal, algebraic, tabular, and graphical forms; and demonstrate knowledge of geometric relationships and corresponding measurement skills. They should be able to apply statistical reasoning in the organization and display of data and in reading tables and graphs. They also should be able to generalize from patterns and examples in the areas of algebra, geometry, and statistics. At this level, they should use correct mathematical language and symbols to communicate mathematical relationships and reasoning processes, and use calculators appropriately to solve problems.
Proficient	Tenth grade students performing at the Proficient level should consistently integrate mathematical concepts and procedures to the solution of more complex problems in the five mathematics standards: Number, Number Sense, and Operation; Measurement; Geometry and Spatial Sense; Patterns, Functions, and Algebra; and Data Analysis and Probability. Tenth graders performing at the Proficient level should demonstrate an understanding of algebraic, statistical and geometric and spatial reasoning. They should be able to perform algebraic operations involving polynomials; justify geometric relationships; and judge and defend the reasonableness of answers as applied to real-world situations. These students should be able to analyze and interpret data in tabular and graphical form; understand and use elements of the function concept in symbolic, graphical, and tabular form; and make conjectures, defend ideas, and give supporting examples.
Advanced	Tenth grade students performing at the Advanced level should consistently demonstrate the integration of procedural and conceptual knowledge and the synthesis of ideas in the five mathematics standards: Number, Number Sense, and Operation; Measurement; Geometry and Spatial Sense; Patterns, Functions, and Algebra; and Data Analysis and Probability. They should apply their knowledge of algebra, geometry, and statistics to solve problems in more advanced areas of continuous discrete mathematics. They should be able to formulate generalizations and create models through probing examples and counterexamples. They should be able to communicate their mathematical reasoning through the clear, concise, and correct use of mathematical symbolism and logical reasoning.

Appendix C: Procedures For Translating Rasch Step Values Into Achievement Levels

obtaining a score of 1 or higher on this item; a student with an achievement level of .99 would have a 50 percent chance of obtaining a score of 2 or higher, and so on. These values are useful checks for the procedure we use to determine the achievement level (in logits) required to have a 2/3 chance of obtaining a given score point or higher.

Steps in the Procedure. Having used BIGSTEPS to produce Rasch step values and Thurstone threshold values for each CR items, we translate those values into Rasch achievement levels using Microsoft Excel. We carry out these calculations by setting up a spreadsheet with incremental values of student achievement (β) in the first column and other values in subsequent columns as described in the steps below, which are based on equations 3 to 6 on page 5 of the plan. The step values are referred to as δ .

1. Column A: Enter values of β from -4 to +4 in increments of .01 (i.e., -4, -3.99, -3.98, etc.)
2. Column B: Numerator value for Step 0. Enter 0 on every row.
3. Column C: Numerator value for Step 1. = exp(value in col. A - δ_{i1}). Copy to remaining rows in this column.
4. Column D: Numerator value for Step 2. = exp (2 * value in col. A - δ_{i1} - δ_{i2}). Copy to remaining rows in this column.
5. Column E: Numerator value for Step 3. = exp(3 * value in col. A - δ_{i1} - δ_{i2} - δ_{i3}). Copy to remaining rows in this column.
6. Column F: Numerator value for Step 4. = exp(4 * value in col. A - δ_{i1} - δ_{i2} - δ_{i3} - δ_{i4}). Copy to remaining rows in this column.
7. Column G: Denominator. =sum(values in col. B - F). Copy to remaining rows in this column.
8. Column H. P value for Step 0. =(value in col. B) / (value in col. G). Copy to remaining rows in this column.
9. Column I. P value for Step 1. =(value in col. C) / (value in col. G). Copy to remaining rows in this column.
10. Column J. P value for Step 2. =(value in col. D) / (value in col. G). Copy to remaining rows in this column.
11. Column K. P value for Step 3. =(value in col. E) / (value in col. G). Copy to remaining rows in this column.
12. Column L. P value for Step 4. =(value in col. F) / (value in col. G). Copy to remaining rows in this column.
13. Column M. Sum of P values. =sum(values in col. H through L). Copy to remaining rows in this column. This is the first check on the accuracy of calculated values. For any given value of β_n , the sum of probabilities should be 1.00. This is a necessary but not sufficient check on the accuracy of calculations to this point.
14. Column N. Probability of 1 or Better. =sum(values in col. I through L). Copy to remaining rows in this column.
15. Column O. Probability of 2 or Better. =sum(values in col. J through L). Copy to remaining rows in this column.
16. Column P. Probability of 3 or Better. =sum(values in col. K and L). Copy to remaining

rows in this column.

17. Column Q. Probability of 4. =(value in col. L).). Copy to remaining rows in this column.
18. For columns N through Q, read down the column of values until you find .5000 or the closest value to .5000. Now read across the row to col. A to find the corresponding value of β_n . This value should correspond to the Thurstone Threshold Value reported in BIGSTEPS. This is the second validity check for the process.
19. Having verified that the $P = .50$ value corresponds to the Thurstone Threshold Value, continue down the column to find the value closest to .6667. Now read across the row to col. A to find the corresponding value of β_n . This value is the student achievement level associated with a $2/3$ chance of obtaining this score or better on this CR item. Use this value in determining the placement of the item in the difficulty-ordered test booklet and in calculating raw score associated with setting a bookmark right after this item/score point.

Sample output is shown in Table C1 on the next page. The full data set is depicted graphically in Figures C1 (P) and C2 (P This Score or Better). In Table C1, we show the first several values of theta (from -4.00 to -3.90) and a series of values from +1.64 to +2.03. These values correspond to the $P = .50$ and $P = .67$ values for step 3. These values are shown in bold in Table C1. There is no value corresponding exactly to .5000. The theta value of 1.64 corresponds to a P of .4977, while the theta value of 1.65 corresponds to a P of .5023. The value .5000 is exactly half way between .4977 and .5023; thus, the theta corresponding to .5000, by interpolation, is 1.645. Because BIGSTEPS reports to only two decimal places, we can conclude that our procedure has matched the Thurstone threshold value ($P = .50$) for step 3. Reading down the same column, we come to the P value of .6676, which is the closest tabled value to .6667. The theta value in this row is 2.03. Thus, we conclude that a student with 2.03 logits of reading achievement would have a $2/3$ chance of receiving a score of 3 or better on item 15.

Table C1 Step Values for Reading Item 15

Theta	Numerator					Denom.	P						P			
	0	1	2	3	4	Sum	0	1	2	3	4	Total	1 or Better	2 or better	3 or better	4
-4.00	1.0000	0.0074	0.0001	0.0000	0.0000	1.0075	0.9926	0.0074	0.0001	0.0000	0.0000	1.0000	0.0074	0.0001	0.0000	0.0000
-3.99	1.0000	0.0075	0.0001	0.0000	0.0000	1.0076	0.9925	0.0075	0.0001	0.0000	0.0000	1.0000	0.0075	0.0001	0.0000	0.0000
-3.98	1.0000	0.0076	0.0001	0.0000	0.0000	1.0077	0.9924	0.0075	0.0001	0.0000	0.0000	1.0000	0.0076	0.0001	0.0000	0.0000
-3.97	1.0000	0.0077	0.0001	0.0000	0.0000	1.0077	0.9923	0.0076	0.0001	0.0000	0.0000	1.0000	0.0077	0.0001	0.0000	0.0000
-3.96	1.0000	0.0078	0.0001	0.0000	0.0000	1.0078	0.9922	0.0077	0.0001	0.0000	0.0000	1.0000	0.0078	0.0001	0.0000	0.0000
-3.95	1.0000	0.0078	0.0001	0.0000	0.0000	1.0079	0.9922	0.0078	0.0001	0.0000	0.0000	1.0000	0.0078	0.0001	0.0000	0.0000
-3.94	1.0000	0.0079	0.0001	0.0000	0.0000	1.0080	0.9921	0.0078	0.0001	0.0000	0.0000	1.0000	0.0079	0.0001	0.0000	0.0000
-3.93	1.0000	0.0080	0.0001	0.0000	0.0000	1.0081	0.9920	0.0079	0.0001	0.0000	0.0000	1.0000	0.0080	0.0001	0.0000	0.0000
-3.92	1.0000	0.0081	0.0001	0.0000	0.0000	1.0081	0.9919	0.0080	0.0001	0.0000	0.0000	1.0000	0.0081	0.0001	0.0000	0.0000
-3.91	1.0000	0.0081	0.0001	0.0000	0.0000	1.0082	0.9919	0.0081	0.0001	0.0000	0.0000	1.0000	0.0081	0.0001	0.0000	0.0000
-3.90	1.0000	0.0082	0.0001	0.0000	0.0000	1.0083	0.9918	0.0082	0.0001	0.0000	0.0000	1.0000	0.0082	0.0001	0.0000	0.0000
1.64	1.0000	2.0959	4.3929	5.1039	2.3164	14.909	0.0671	0.1406	0.2946	0.3423	0.1554	1.0000	0.9329	0.7923	0.4977	0.1554
1.65	1.0000	2.1170	4.4817	5.2593	2.4109	15.268	0.0655	0.1386	0.2935	0.3444	0.1579	1.0000	0.9345	0.7959	0.5023	0.1579
1.66	1.0000	2.1383	4.5722	5.4195	2.5093	15.639	0.0639	0.1367	0.2924	0.3465	0.1604	1.0000	0.9361	0.7993	0.5070	0.1604
1.67	1.0000	2.1598	4.6646	5.5845	2.6117	16.020	0.0624	0.1348	0.2912	0.3486	0.1630	1.0000	0.9376	0.8028	0.5116	0.1630
1.68	1.0000	2.1815	4.7588	5.7546	2.7183	16.413	0.0609	0.1329	0.2899	0.3506	0.1656	1.0000	0.9391	0.8062	0.5162	0.1656
1.69	1.0000	2.2034	4.8550	5.9299	2.8292	16.817	0.0595	0.1310	0.2887	0.3526	0.1682	1.0000	0.9405	0.8095	0.5208	0.1682
1.70	1.0000	2.2255	4.9530	6.1104	2.9447	17.233	0.0580	0.1291	0.2874	0.3546	0.1709	1.0000	0.9420	0.8128	0.5254	0.1709
1.71	1.0000	2.2479	5.0531	6.2965	3.0649	17.662	0.0566	0.1273	0.2861	0.3565	0.1735	1.0000	0.9434	0.8161	0.5300	0.1735
1.72	1.0000	2.2705	5.1552	6.4883	3.1899	18.103	0.0552	0.1254	0.2848	0.3584	0.1762	1.0000	0.9448	0.8193	0.5346	0.1762
1.73	1.0000	2.2933	5.2593	6.6859	3.3201	18.558	0.0539	0.1236	0.2834	0.3603	0.1789	1.0000	0.9461	0.8225	0.5392	0.1789
1.74	1.0000	2.3164	5.3656	6.8895	3.4556	19.027	0.0526	0.1217	0.2820	0.3621	0.1816	1.0000	0.9474	0.8257	0.5437	0.1816
1.75	1.0000	2.3396	5.4739	7.0993	3.5966	19.509	0.0513	0.1199	0.2806	0.3639	0.1844	1.0000	0.9487	0.8288	0.5482	0.1844
1.76	1.0000	2.3632	5.5845	7.3155	3.7434	20.006	0.0500	0.1181	0.2791	0.3657	0.1871	1.0000	0.9500	0.8319	0.5528	0.1871
1.77	1.0000	2.3869	5.6973	7.5383	3.8962	20.518	0.0487	0.1163	0.2777	0.3674	0.1899	1.0000	0.9513	0.8349	0.5573	0.1899
1.78	1.0000	2.4109	5.8124	7.7679	4.0552	21.046	0.0475	0.1146	0.2762	0.3691	0.1927	1.0000	0.9525	0.8379	0.5618	0.1927
1.79	1.0000	2.4351	5.9299	8.0045	4.2207	21.590	0.0463	0.1128	0.2747	0.3707	0.1955	1.0000	0.9537	0.8409	0.5662	0.1955
1.80	1.0000	2.4596	6.0496	8.2482	4.3929	22.150	0.0451	0.1110	0.2731	0.3724	0.1983	1.0000	0.9549	0.8438	0.5707	0.1983
1.81	1.0000	2.4843	6.1719	8.4994	4.5722	22.727	0.0440	0.1093	0.2716	0.3740	0.2012	1.0000	0.9560	0.8467	0.5751	0.2012
1.82	1.0000	2.5093	6.2965	8.7583	4.7588	23.322	0.0429	0.1076	0.2700	0.3755	0.2040	1.0000	0.9571	0.8495	0.5796	0.2040
1.83	1.0000	2.5345	6.4237	9.0250	4.9530	23.936	0.0418	0.1059	0.2684	0.3770	0.2069	1.0000	0.9582	0.8523	0.5840	0.2069
1.84	1.0000	2.5600	6.5535	9.2999	5.1552	24.568	0.0407	0.1042	0.2667	0.3785	0.2098	1.0000	0.9593	0.8551	0.5884	0.2098
1.85	1.0000	2.5857	6.6859	9.5831	5.3656	25.220	0.0397	0.1025	0.2651	0.3800	0.2127	1.0000	0.9603	0.8578	0.5927	0.2127
1.86	1.0000	2.6117	6.8210	9.8749	5.5845	25.892	0.0386	0.1009	0.2634	0.3814	0.2157	1.0000	0.9614	0.8605	0.5971	0.2157
1.87	1.0000	2.6379	6.9588	10.175	5.8124	26.584	0.0376	0.0992	0.2618	0.3828	0.2186	1.0000	0.9624	0.8632	0.6014	0.2186
1.88	1.0000	2.6645	7.0993	10.485	6.0496	27.299	0.0366	0.0976	0.2601	0.3841	0.2216	1.0000	0.9634	0.8658	0.6057	0.2216
1.89	1.0000	2.6912	7.2427	10.804	6.2965	28.035	0.0357	0.0960	0.2583	0.3854	0.2246	1.0000	0.9643	0.8683	0.6100	0.2246
1.90	1.0000	2.7183	7.3891	11.134	6.5535	28.794	0.0347	0.0944	0.2566	0.3867	0.2276	1.0000	0.9653	0.8709	0.6143	0.2276
1.91	1.0000	2.7456	7.5383	11.473	6.8210	29.577	0.0338	0.0928	0.2549	0.3879	0.2306	1.0000	0.9662	0.8734	0.6185	0.2306
1.92	1.0000	2.7732	7.6906	11.822	7.0993	30.385	0.0329	0.0913	0.2531	0.3891	0.2336	1.0000	0.9671	0.8758	0.6227	0.2336
1.93	1.0000	2.8011	7.8460	12.182	7.3891	31.218	0.0320	0.0897	0.2513	0.3902	0.2367	1.0000	0.9680	0.8782	0.6269	0.2367
1.94	1.0000	2.8292	8.0045	12.553	7.6906	32.077	0.0312	0.0882	0.2495	0.3913	0.2397	1.0000	0.9688	0.8806	0.6311	0.2397
1.95	1.0000	2.8577	8.1662	12.935	8.0045	32.964	0.0303	0.0867	0.2477	0.3924	0.2428	1.0000	0.9697	0.8830	0.6352	0.2428
1.96	1.0000	2.8864	8.3311	13.329	8.3311	33.878	0.0295	0.0852	0.2459	0.3935	0.2459	1.0000	0.9705	0.8853	0.6394	0.2459
1.97	1.0000	2.9154	8.4994	13.735	8.6711	34.821	0.0287	0.0837	0.2441	0.3945	0.2490	1.0000	0.9713	0.8876	0.6435	0.2490
1.98	1.0000	2.9447	8.6711	14.154	9.0250	35.794	0.0279	0.0823	0.2422	0.3954	0.2521	1.0000	0.9721	0.8898	0.6476	0.2521
1.99	1.0000	2.9743	8.8463	14.585	9.3933	36.799	0.0272	0.0808	0.2404	0.3963	0.2553	1.0000	0.9728	0.8920	0.6516	0.2553
2.00	1.0000	3.0042	9.0250	15.029	9.7767	37.835	0.0264	0.0794	0.2385	0.3972	0.2584	1.0000	0.9736	0.8942	0.6556	0.2584
2.01	1.0000	3.0344	9.2073	15.487	10.175	38.904	0.0257	0.0780	0.2367	0.3981	0.2616	1.0000	0.9743	0.8963	0.6596	0.2616
2.02	1.0000	3.0649	9.3933	15.958	10.591	40.007	0.0250	0.0766	0.2348	0.3989	0.2647	1.0000	0.9750	0.8984	0.6636	0.2647
2.03	1.0000	3.0957	9.5831	16.444	11.023	41.146	0.0243	0.0752	0.2329	0.3997	0.2679	1.0000	0.9757	0.9005	0.6676	0.2679

Reading Item 15

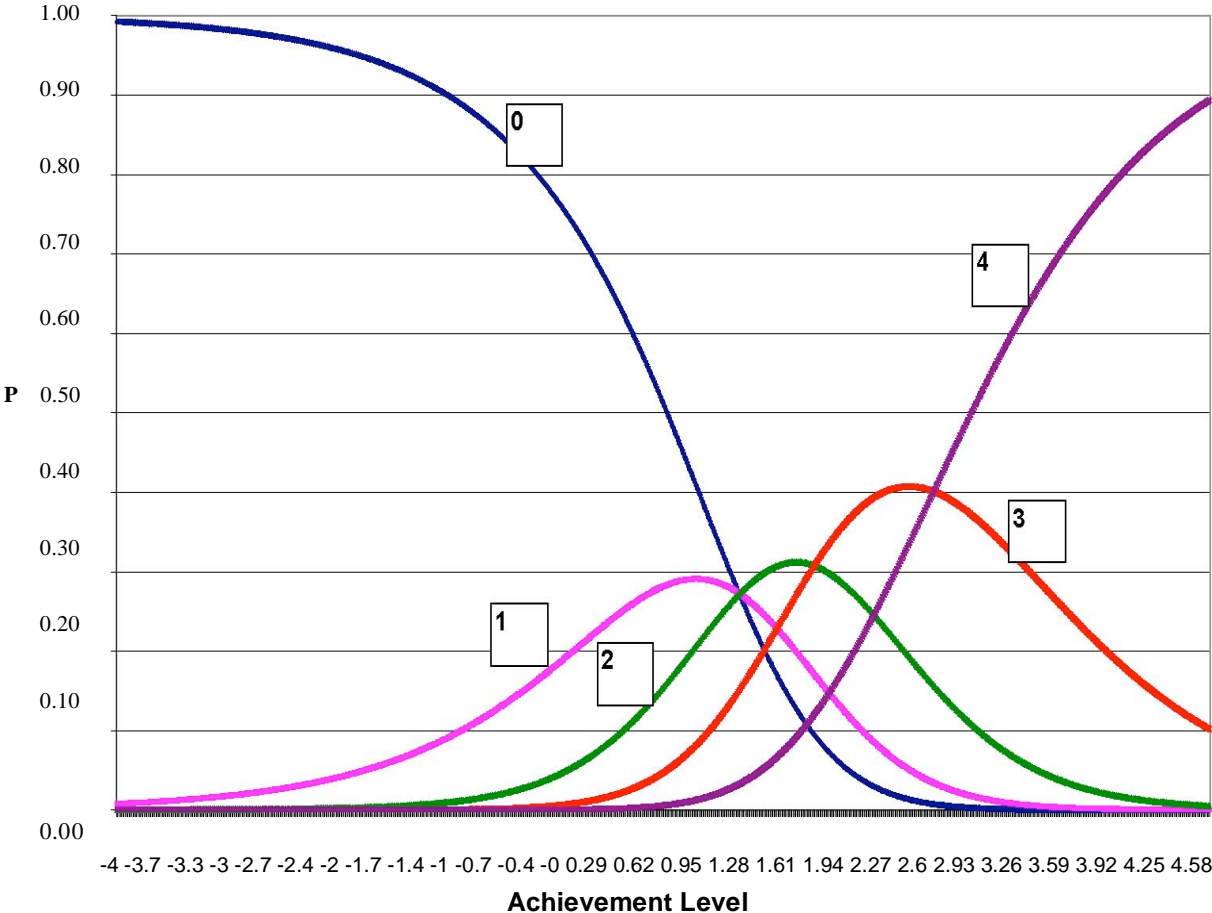


Figure C1. Probability of obtaining score points 0 – 4 as a function of Achievement Level

Reading Item 15

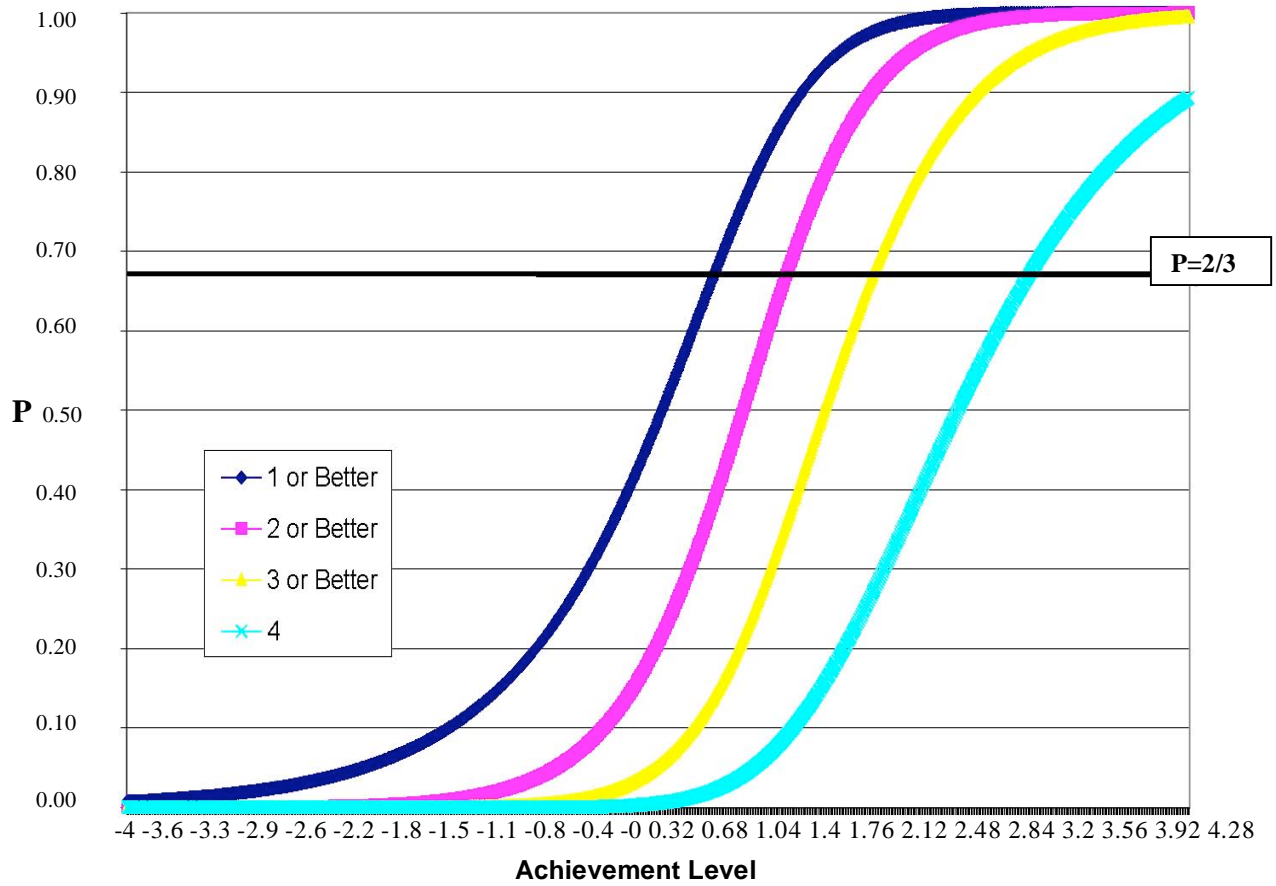


Figure C2. Probability of obtaining a given score point or better as a function of Achievement Level

Appendix D: Report to the State Board of Education

**Office of Assessment
May 10, 2004**

Table of Contents

Topic	Page Number
Preparation for standard setting	3 [85]
Standard setting committees	3 [85]
Performance level descriptors	3 [85]
Procedures	4 [86]
Reading results	5 [87]
Mathematics results	6 [88]
Reaction of oversight committees	7 [89]
Appendices	
DA – List of standard setting committee members	8 [89]
DB – Performance level descriptors	11 [92]
DC – Committee recommended cut scores and percentages of students	14 [95]

Standard setting report to the State Board of Education May 2004

Preparation for standard setting. Members of the Department of Education, Office of Assessment and their Technical Advisory Committee (TAC) spent considerable time over several meetings talking about plans for standard setting for the Ohio Graduation Tests. Two critical topics during those discussions were committee membership and the procedures that would be used during standard setting.

Standard setting committees. As required by Senate Bill 1 and as recommended by the TAC, Office of Assessment staff involved the following groups of people in the standard setting committees:

Active teachers

Parents

Business and community representatives

All three groups were represented in the standard setting committees.

Because the achievement standards have to do with earning an Ohio diploma, the TAC strongly recommended that the teaching members of the committee should reflect not just those teaching the subject area at the high school level; teachers of the specific subject area at lower grades (especially middle school grades), teachers with a variety of years of teaching experience and teachers of other subject areas participated in the standard setting committees.

The lists of committee members are included in Appendix A.

Performance level descriptors. Curriculum consultants worked with Assessment staff to look over samples of what other states and organizations have used to describe a variety of performance or achievement levels. After several drafts, we had performance level descriptors for both reading and mathematics that would support the designations required by House Bill 3 and by No Child Left Behind: Limited, Basic, Proficient, Accelerated and Advanced. The Fairness/Sensitivity Committee reacted to the draft descriptors in both reading and mathematics; the specific draft descriptor was shared with the specific content advisory committee. Both groups made changes to the drafts.

Each of the standard setting committees used the appropriate performance level descriptors to help committee members understand the differences between the various performance levels. Committee members were told to think about the kinds of skills that a student who was just barely at a performance level would have. They were asked to think about the just barely proficient student first. After that, committees focused on the just barely basic students, then the just barely accelerated students and finally the just barely advanced students.

The performance level descriptors for reading and mathematics are included in Appendix B.

Procedures. Measurement Incorporated, the contractor chosen through competitively bidding, submitted their plans for standard setting using the bookmark method. The Department of Education and the Technical Advisory Committee agreed to this method for standard setting on the Ohio Graduation Tests.

The bookmark method of standard setting is frequently used when there is a combination of item formats (multiple choice and open-ended items) used within a test. Ohio used this same method when we set the basic achievement level for the 4th grade reading proficiency test in the fall of 2001.

As with all standard setting activities, the committees began their respective work by taking the appropriate Ohio Graduation Test. The activity of taking the test helped committee members to learn first-hand about the test that students took in March 2004. Committee members used answer keys and scoring rubrics with sample student work to score their own multiple-choice and open-ended responses.

The next activity was for the committee members to understand the performance level descriptors levels. Much discussion occurred around what it meant in a subject for the student performance to be just barely basic, just barely proficient, just barely accelerated or just barely advanced. The purpose of the extended group discussions was to make sure that committee members had a clear understanding of the achievement levels. Members were instructed that they would be recommending the proficient score first since legislation (original legislation in 1987 and recent legislation in 2001) connected proficient performance on graduation tests with meeting requirements to earn an Ohio diploma.

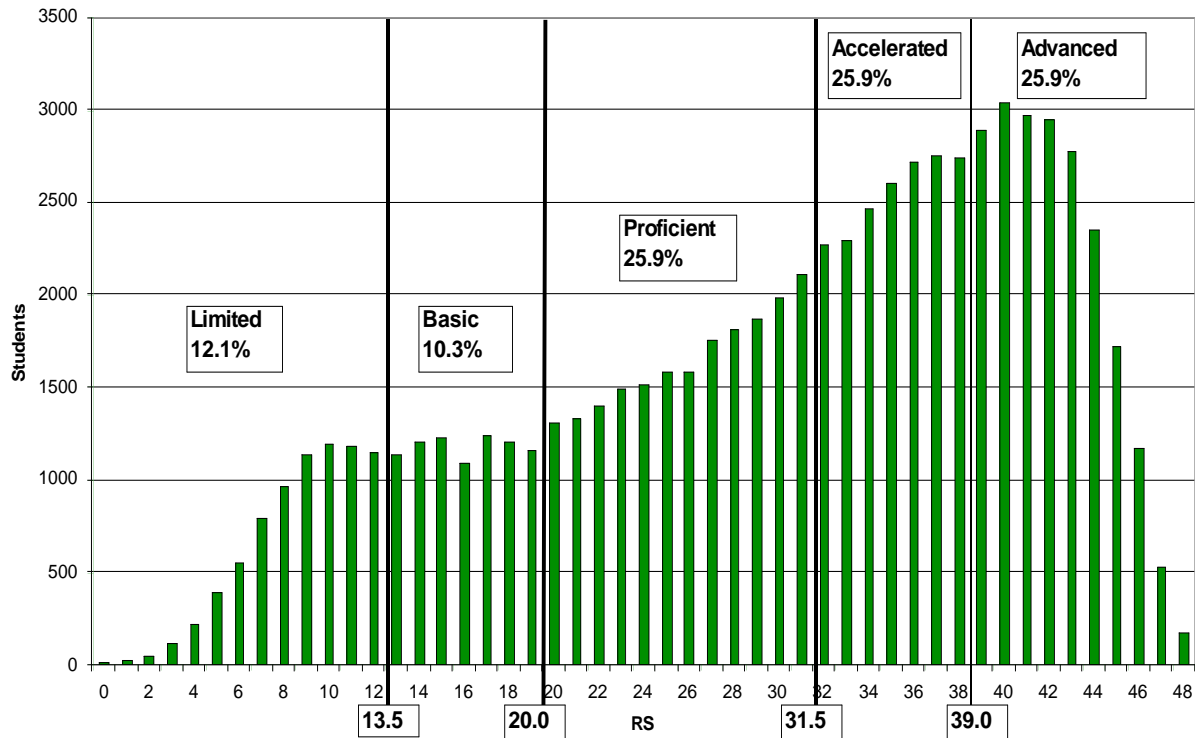
Members were then trained in the bookmark procedure which required them to place their vote (mark) within a difficulty-ordered test booklet. The test items they had previously seen in the same order that students had seen in the March 2004 administration had been reordered into order from easiest to hardest. The committees practiced on a small number of items before they then began the three rounds of standard setting voting. Before each round, committee members indicated their understanding of the task and their readiness to proceed.

After members received the results of round one voting, they received impact data of how all students did on the specific test during the March 2004 administration. In addition, data of how subgroups did on the March 2004 administration was also shared. Members were divided into small groups for discussion of results after rounds one and two.

Results of the final recommended cut scores and the percentages of students within each achievement level by race and gender are shared in Appendix C.

What the Reading standard setting committee recommended and how a large (approximately 75,000) representative sample of students did on the March 2004 administration:

Reading Raw Score Distribution With Cut Scores and Percentages of Students in Each Category



Final Cut Score Recommendations and Percentages of Students in Each Category

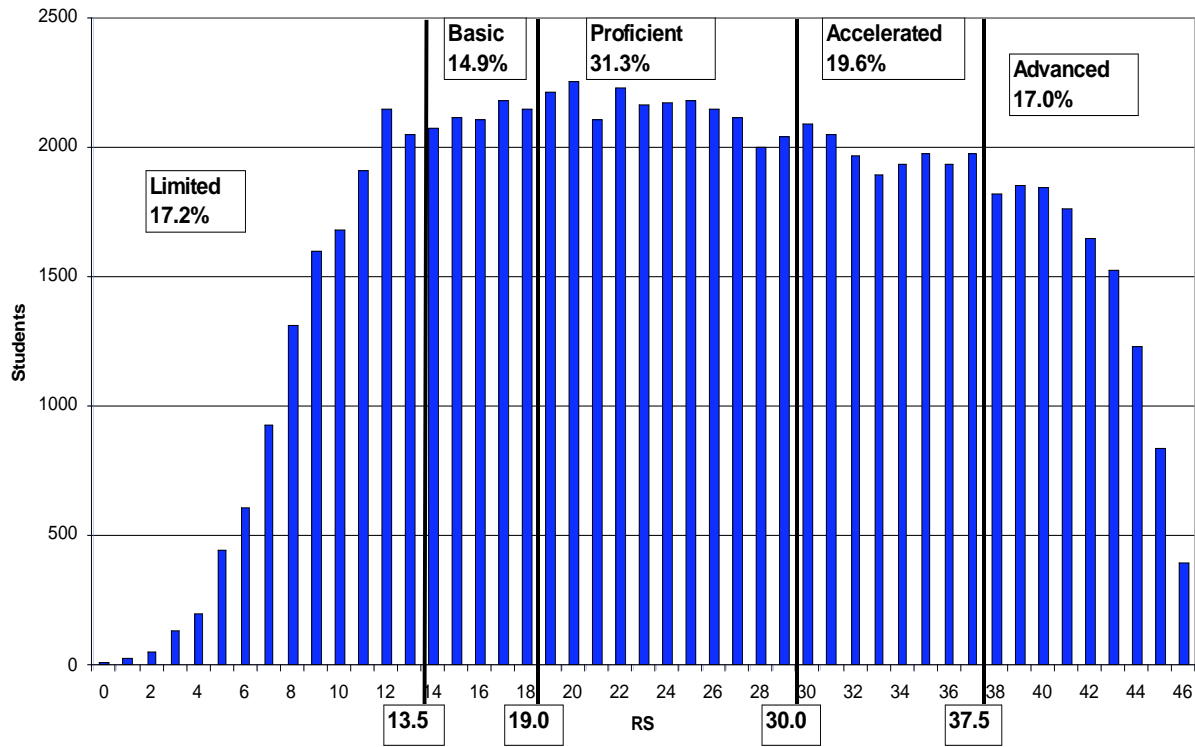
	READING		
	Cut Score (Out of 48)	Percent of Total Points	Percent in Category
Limited	--	---	12.1
Basic	13.5	28%	10.3
Proficient	20.0	42%	25.9
Accelerated	31.5	66%	25.9
Advanced	39.0	81%	25.9

Summary of Reading Recommendation at Proficient or above:

A total of 77.7 percent of students from the March 2004 administration would be identified as proficient or higher.

What the Mathematics standard setting committee recommended and how a large (approximately 75,000) representative sample of students did on the March 2004 administration:

Mathematics Raw Score Distribution With Cut Scores and Percentages of Students in Each Category



	MATHEMATICS		
	Cut Score (Out of 46)	Percent of Total Points	Percent in Category
Limited	---	---	17.2
Basic	13.5	29%	14.9
Proficient	19.0	41%	31.3
Accelerated	30.0	65%	19.6
Advanced	37.5	82%	17.0

Summary of Mathematics Recommendation at Proficient or above:

A total of 67.9% percent of students from the March 2004 administration would be identified as proficient or higher.

Appendix DA: List of Standard Setting Committee Members

LIST OF COMMITTEE MEMBERS – READING

	First Name	Last name	Position/Occupation	District/Business/Organization
1	Michael	Bader	Administrator	Sylvania City Schools
2	Maria	Giles	Community	Lorain
3	Chris	Goodman	Teacher	US Grant Career Center
4	Kristina	Harris	Teacher	Wilmington City Schools
5	Karianne	Havlicek	Teacher	Southern Local Schools
6	Ellen	Isbell	Teacher	Centerville City Schools
7	Cheryl	King	Curriculum Specialist	Toledo Public Schools
8	Steve	Kucinski	Teacher	Dublin City Schools
9	Charles	Landry	Teacher	Akron City Schools
10	Suzanne	MacLean	Teacher	Delaware Area Career Center
11	Jim	Nowack	Business	Eaton Corporation- Perrysburg St. Charles Preparatory High School
12	Scott	Pharion	Teacher	Crestwood Local Schools
13	Jennifer	Powers	Teacher	University of Toledo
14	Alexa	Sandman	Higher Education	Toledo School for the Arts
15	Kathy	Scott	Teacher	Tri-Valley Local Schools
16	Douglas	Spade	Superintendent	North Central Local Schools
17	Ben	Spieldenner	Teacher	Cleveland Municipal Schools
18	Rita	Stroempl	Teacher	Indian Creek Local Schools
19	Cynthia	Thomas	Teacher	Grove City
20	Michael	Uhrin	Business	Kettering City Schools
21	Jan	Wagoner	Teacher	Warrensville Heights Schools
22	Virgil	Walker	Teacher	Napoleon Area City Schools
23	Barbara	Ward-Bovee	Administrator	Bexley City Schools
24	Janice	Yoder	Teacher	

Gender Breakdown

Female	14
Male	10

Position Breakdown

Curriculum Specialist	1
Administrator	3
Teacher	16
Higher Education	1
Business/Community Representative	3

Ethnic Breakdown

African- American	3
Caucasian	20
Hispanic	1

LIST OF COMMITTEE MEMBERS – MATHEMATICS

	First Name	Last name	Position/Occupation	District/Business/Organization
1	Idrissa	Aidara	Teacher	Cleveland Municipal Schools Madison Local Schools, Lake
2	Keith	Brainard	Teacher	County
3	Lisa	Bricker	Teacher	Greene County JVSD
4	Carole	Bridges	Teacher	Forest Hills Local Schools
5	Faith	Bruner	Teacher	Bishop Ready High School
6	Ayhan	Caputlu	Teacher	Horizon Science Academy
7	Carol	Corns	Teacher	Zanesville City Schools
8	Ann	Farrell	University Professor	Wright State University
9	Margo	Fisher	Teacher	Tiffin City Schools
10	Brandon	Grimes	Teacher	Hardin Local Schools
11	Julie	Henricks	Teacher	South-Western City Schools
12	Charniece	Holmes	Parent	Shaker Heights
13	Rosanna	Johnson	Teacher School Board	Jonathan Alder Schools
14	Johnnie	Kimberlin	Member	Jefferson Township Local
15	Bob	Knight	Business	Burgess and Niple, Cincinnati
16	Heather	Krysty	Teacher	Stark County Area Vocational SD
17	Judy	Lamoreux	Teacher	Wilmington City Schools
18	Holly	Lavender	Teacher	Liberty Union Thurston Schools
19	Kimya	Moyo	Curriculum Specialist	Cincinnati City Schools
20	Steven	Proehl	Teacher	Western Local Schools
21	Elizabeth	Reed	Community	City of Columbus
22	Marty	Schloegl	Administrator	Lake Local Schools
23	Stewart	Lisa	Business	Management Recruiters, Columbus

Gender Breakdown

Female	15
Male	8

Position Breakdown

Curriculum Specialist	1
Administrator	1
Teacher	15
Higher Education	1
Business/Community	4
School Board member	1

Ethnic Breakdown

African- American	3
Caucasian	17
Asian/Pacific Islander	1
Other	2

Appendix DB: Performance Level Descriptors

**Ohio Graduation Tests
Performance Level Descriptors for Reading
April 2004**

Limited	Tenth grade students performing at the Limited level demonstrate skill and understanding of reading below the performance required to reach the Basic level.
Basic	Tenth grade students performing at the Basic level can generally define unfamiliar or complex words through contextual clues and can determine resources to define or understand the more complex words. They can demonstrate some understanding, and are able to make some interpretations and judgments of textual information.
Proficient	Tenth grade students performing at the Proficient level use their basic understanding of word structure, context clues and text structures to determine the meaning of unfamiliar or complex words. They can typically show an overall understanding of and make evaluative judgments of textual information. Students are generally able to identify and explain the various ways authors may influence text and assess the appropriateness of provided information.
Accelerated	Tenth grade students performing at the Accelerated level use their understanding of word structure, context clues and text structures to determine the meaning of unfamiliar or complex words. They show an overall understanding of and make evaluative and analytical judgments of textual information. Students are able to explain and analyze the various ways authors may influence text and assess the appropriateness of provided information.
Advanced	Tenth grade students performing at the Advanced level apply their understanding of word structure, context clues and text structures to determine the meaning of unfamiliar or complex words. They consistently demonstrate a firm grasp of the methods used by authors to affect the meaning and appropriateness of text. They are able to explain and analyze the various ways authors may influence text and assess the appropriateness of provided information.

**Ohio Graduation Tests
Performance Level Descriptors for Mathematics
April 2004**

Limited	Tenth grade students performing at the Limited level demonstrate skill and understanding of mathematics below the performance required to reach the Basic level.
Basic	Tenth grade students performing at the Basic level recall and recognize mathematical concepts, terms and properties. Students typically carry out routine procedures, such as reading graphs, performing specified computations and solving simple equations. Students solve problems for which the method or solution is easily recognized and straightforward.
Proficient	Tenth grade students performing at the Proficient level apply mathematical concepts, terms and properties to problem situations. Most times, students can solve problems with two or more steps or decision points. They usually make appropriate decisions about what to do and can use informal reasoning and problem-solving strategies. Students typically can interpret or provide a visual or symbolic representation to match a problem situation and purpose. Students communicate mathematical thinking and solutions using a combination of informal and mathematical language.
Accelerated	Tenth grade students performing at the Accelerated level recognize similarities and differences between various mathematical concepts, properties and procedures. They consistently bring together skills and knowledge from various concepts and domains in mathematics to solve problems involving multiple steps and decision points. Students use informal and some formal reasoning to evaluate and justify the reasonableness of a solution. They communicate mathematical thinking and solutions in a clear and concise manner.
Advanced	Tenth grade students performing at the Advanced level routinely identify and connect fundamental mathematical concepts, properties and procedures to more complex and novel problem situations. Students typically demonstrate more abstract and sophisticated thinking in their analysis of, approach to and solutions for problems. They provide a formal mathematical justification using precise mathematical language and notations. Students consistently demonstrate deep knowledge and skills across the standards.

Appendix DC: Committee Recommended Cut Scores and Percentages Race and Gender

READING

Committee Recommended Cut Scores and Percentages of Students in Each Category by Race and Gender

	Cut Score (% Out of 48)	Group	Percent in Category
Limited	- - -	Am. Indian	21.19
		Asian/PI	8.61
		Black/AA	23.88
		Hispanic	22.08
		White	9.38
		Multi	14.73
		Other	22.84
		Female	7.28
		Male	16.58
		Basic	13.5 (28%)
Asian/PI	7.12		
Black/AA	17.73		
Hispanic	18.30		
White	8.64		
Multi	13.07		
Other	12.69		
Female	8.68		
Male	11.90		
Proficient	20 (42%)		
		Asian/PI	20.67
		Black/AA	33.07
		Hispanic	28.12
		White	24.43
		Multi	27.07
		Other	29.44
		Female	26.32
		Male	25.41
		Accelerated	31.5 (66%)
Asian/PI	25.49		
Black/AA	17.35		
Hispanic	20.29		
White	27.74		
Multi	23.55		
Other	21.32		
Female	27.37		
Male	24.48		

READING

Committee Recommended Cut Scores and Percentages of Students in Each Category by Race and Gender (continued)

	Cut Score (Out of 48)	Group	Percent in Category
Advanced	39 (81%)	Am. Indian	11.02
		Asian/PI	38.12
		Black/AA	7.96
		Hispanic	11.21
		White	29.81
		Multi	21.58
		Other	13.71
		Female	30.35
		Male	21.63

MATHEMATICS

Committee Recommended Cut Scores and Percentages of Students in Each Category by Race and Gender

	Cut Score (% Out of 46)	Group	Percent in Category
Limited	- - -	Am. Indian	28.78
		Asian/PI	8.06
		Black/AA	36.43
		Hispanic	30.82
		White	13.20
		Multi	23.40
		Other	28.92
		Female	14.40
Male	19.86		
Basic	13.5 (29%)	Am. Indian	10.79
		Asian/PI	10.22
		Black/AA	24.65
		Hispanic	22.08
		White	12.93
		Multi	16.49
		Other	15.20
		Female	15.60
Male	14.17		
Proficient	19 (41%)	Am. Indian	34.53
		Asian/PI	24.29
		Black/AA	28.59
		Hispanic	29.62
		White	31.95
		Multi	30.43
		Other	35.29
		Female	34.04
Male	28.61		
Accelerated	30 (65%)	Am. Indian	12.95
		Asian/PI	21.45
		Black/AA	7.82
		Hispanic	10.81
		White	22.19
		Multi	16.49
		Other	7.84
		Female	20.19
Male	19.14		

**Committee Recommended Cut Scores and Percentages of
Students in Each Category by Race and Gender**
(continued)

Cut Score		Group	Percent in
(% Out of 46)			Category
Advanced	37.5 (82%)	Am. Indian	12.95
		Asian/PI	35.98
		Black/AA	2.51
		Hispanic	6.67
		White	19.72
		Multi	13.19
		Other	12.75
		Female	15.77
		Male	18.22