



**Technical Report on the SAT/NAEP Grade 12 Preliminary Comparability Study—
Reading¹**

Mary J. Pitoniak

Clyde Reese

Richard J. Tannenbaum

Educational Testing Service

Submitted to the College Board on April 7, 2008

¹ The authors are grateful to Jennifer Vivier, Nicole Beaulieu, and Barbara Elkins for their assistance in planning for and conducting the meeting, and to Janice Lukas and Mei-Jang Lin for their reviews of an earlier version of this report.

Contents

List of Tables.....	iii
List of Figures.....	iv
List of Appendices.....	v
Background.....	1
Purpose.....	1
Comparability.....	1
Design.....	1
Panelists.....	1
Descriptions of Content Areas.....	2
SAT Critical Reading.....	2
NAEP Grade 12 Reading.....	2
Ratings.....	2
Content-Area Ratings.....	3
Overall Ratings.....	3
Meeting Procedures.....	3
Welcome, Introductions, and Background Overview.....	3
Content Overviews.....	4
Training in Content-Area Ratings.....	4
First Evaluation Form.....	5
Round 1 Ratings.....	5
Feedback and Discussion.....	5
Round 2 Ratings.....	5
Overall Ratings.....	5
Final Evaluation.....	5
Debriefing.....	6
Results.....	6
Content-Level Ratings.....	6
Summary-Level Results.....	6
Panelist-Level Results.....	6
Overall Ratings.....	7
Summary-Level Results.....	7
Panelist-Level Results.....	7

Panelist Feedback.....	8
Initial Evaluation	8
Final Evaluation.....	8
Discussion and Conclusions	9
References.....	10

List of Tables

Table		Page
Table 1	Panelist Characteristics	11
Table 2	Comparison of Panelist and Committee Characteristics	11
Table 3	Number of Performance Characteristics Descriptors for SAT Critical Reading Section	12
Table 4	Number of Preliminary Achievement Level Descriptions for NAEP 12th Grade Reading Assessment	12
Table 5	Meeting Agenda	13
Table 6	Sample Display of Data from Round 1 Ratings	14
Table 7	Ratings for Rounds 1 and 2.....	15
Table 8	Number of Changes in Ratings from Round 1 to Round 2 by Panelist	16
Table 9	Descriptors Reviewed and Ratings Revised from Round 1 to Round 2.....	17
Table 10	Summary of Overall Ratings	18
Table 11	Overall Ratings by Panelist	18
Table 12	Summary of Responses to Initial Evaluation Form.....	19
Table 13	Responses to Initial Evaluation Form by Panelist	20
Table 14	Summary of Responses to Final Evaluation Form	21
Table 15	Responses to Final Evaluation Form by Panelist.....	22
Table 16	Responses to Open-Ended Question on Final Evaluation Form by Panelist.....	23

List of Figures

Figure	Page
Figure 1 Round 1 Ratings by Content Area.....	24
Figure 2 Round 2 Ratings by Content Area.....	25
Figure 3 Changes in Ratings Over Rounds by Content Area.....	26
Figure 4 Round 1 Ratings by Panelist.....	27
Figure 5 Round 2 Ratings by Panelist.....	28
Figure 6 Changes in Ratings Over Rounds by Panelist.....	29

List of Appendices

Appendix A	SAT Critical Reading Performance Category Descriptors
Appendix B	NAEP 12th Grade Reading Preliminary Achievement Level Descriptions
Appendix C	National Assessment Governing Board PowerPoint Presentation on Preparedness
Appendix D	SAT Critical Reading Content Overview PowerPoint Presentation
Appendix E	NAEP 12th Grade Reading Content Overview PowerPoint Presentation
Appendix F	Rating Process Training PowerPoint Presentation
Appendix G	Content-Area Rating Form
Appendix H	Initial Evaluation/Ready-to-Proceed Form
Appendix I	Overall Rating Form
Appendix J	Final Evaluation Form

Background

Purpose

The purpose of this study was to examine the overall comparability of the SAT critical reading section and the NAEP grade 12 reading assessment. The scope of this preliminary examination of comparability was to determine feasibility prior to funding a more extensive and rigorous alignment study. If NAEP assessments and the SAT sections are judged to be comparable at a content specifications level, it may be reasonable to move forward with an extensive study that examines alignment at the item and scale levels.

Investigation of the alignment of NAEP and the SAT is part of an effort to determine ways to evaluate how grade 12 NAEP can be used to report students' preparedness for post-secondary activities. Content alignment studies are recommended as a first step to other studies that relate performance on another assessment to NAEP.

Comparability

A key question, however, is what constitutes comparability—how comparability is to be defined, conceptually and operationally. Issues of comparability, to date, primarily have focused on gathering judgment-based evidence regarding how well tests align with content standards (e.g., Bhola, Impara, & Buckendahl, 2003; Webb, 1997; Webb, Herman, & Webb, 2007). Similar methods and criteria for alignment, however, may be applied to a test-to-test alignment investigation. In this context, two different levels of comparability may be considered.

- The first, at the test content specification level, addresses the basic question: “Do the two tests measure the same content areas within the given subject?”
- The second, at the level of the individual items, focuses on the cognitive demands of the items: “Do the items on the two tests require the same depth of content understanding?”

For this preliminary study, the first question is examined—the focus is on NAEP and SAT frameworks or specifications. On the basis of the results, the College Board and the National Assessment Governing Board may proceed to a more extensive study to address the second question.

Design

Panelists

Evaluations of comparability were based on expert judgment. A panel of reading experts drawn from the past or current SAT and NAEP standing content committees was assembled. Shown in table 1 are the characteristics of the panel members. Table 2 compares the characteristics of the panel members with those of the NAEP and SAT standing committees. The proportion of panel members in each occupational setting (K-12 vs. post-secondary) is relatively evenly distributed, which is in line with both of the standing committees. In terms of region, the only region not represented on the panel

was the northeast. It should be noted, however, one of the panelists was from Delaware but was coded as being from the south as per Census region guidelines. Therefore, the panel is somewhat more diverse geographically than the region data suggest.

Descriptions of Content Areas

SAT Critical Reading

For the SAT Critical Reading section, performance characteristic descriptors were used as an outline for the content in the section. As shown in table 3, there are five content areas in the assessment: (A) Determining the Meaning of Words, (B) Understanding Literary Elements, (C) Organization and Ideas, (D) Author’s Craft, and (E) Reasoning and Inferencing. There are in turn 75 descriptors spread across these content areas.

These descriptors are “detailed explanations of what students within specific SAT score intervals know and can do as indicated by their performance on the SAT” (College Board, 2007). They were derived by looking at the types of items that students within given intervals (such as 200 to 290, 300 to 390, and so on) answered correctly, and then developing descriptions of the knowledge skills needed to answer them. For the purpose of this study, the descriptors were grouped by content area, with descriptors for lower score levels preceding those from higher ones (numeric score bands were not identified). The SAT Critical Reading performance category descriptors can be found in Appendix A.

NAEP Grade 12 Reading

For the NAEP Grade 12 Reading Assessment, preliminary achievement-level descriptions served as the list of content covered by the test. The descriptions were broken down by the two types of passages—literary and informational—as shown in table 4.

The descriptions were designed by the framework development panels as detailed lists of reading skills and abilities students should have to demonstrate performance at each level of achievement. They were developed to inform passage selection and item development to help ensure coverage of appropriate skill levels for each grade, and to provide a “checklist” for use in the development of narrative descriptions that are used in setting achievement levels and reporting NAEP results. The NAEP preliminary achievement-level descriptions are included in Appendix B.

It should be noted that there were no grade 12 specific preliminary achievement-level descriptions for the vocabulary subscale. Instead, panelists looked at sample vocabulary items to ascertain the knowledge and skills needed to answer them.

Ratings

Test-level comparability addresses the extent to which the content areas covered by the SAT critical reading test are addressed by the content areas in the NAEP grade 12 reading assessment. Panelists were asked to make two basic kinds of ratings: (1) content-area ratings and (2) overall ratings.

Content-Area Ratings

After reviewing the SAT and NAEP descriptors² the panelists were asked to do the following:

- Refer to the NAEP descriptors to see whether the SAT descriptor is covered by NAEP. For example, if an SAT critical reading performance characteristic descriptor states that students should be able to “analyze and compare concepts across text,” is there a preliminary NAEP grade 12 reading achievement-level descriptor that covers the same content? This judgment was made using a “Yes” or “No” rating.
- For those “matched” content categories for which the panelist responded “yes,” the panelist was asked to do the following:
 - Indicate the strength of the NAEP-SAT comparability for that SAT descriptor. This judgment was made using a 3-point rating scale, ranging from (1) weak to (2) moderate to (3) strong.
 - Identify the specific NAEP descriptor(s) that cover the SAT descriptor.

Overall Ratings

The panelists were also asked to make three overall ratings.

- 1) The first was a judgment of whether the content of the two tests is comparable, using a 4-point scale ranging from Strongly Agree to Strongly Disagree.
- 2) The second was a judgment of whether the overall breadth of reading skills on the two tests is comparable, using a 4-point scale ranging from Strongly Agree to Strongly Disagree.
- 3) The final was a Yes or No judgment of whether, based on the panelist’s evaluation of comparability, he or she thinks that a follow-up study comparing items on each test is justified.

Meeting Procedures

The meeting was conducted on Monday, January 28, 2008, with five panelists, five staff from ETS (facilitators and test development staff), one representative from the College Board, and one representative from the National Assessment Governing Board. The agenda for the meeting appears in table 5.

Welcome, Introductions, and Background Overview

Following a general introduction and welcome, all participants were asked to introduce themselves, and the agenda for the day was reviewed.

² For the purpose of this report, the term *descriptor* will be used as shorthand for both the SAT performance characteristic descriptors and NAEP preliminary achievement-level descriptions.

A representative for the National Assessment Governing Board then provided background for the study; the Governing Board presentation is included in Appendix C. A representative from the College Board also spoke and provided an overview of how the performance characteristic descriptors were developed.

Content Overviews

Next, one ETS test development staff member provided an overview of the SAT assessment, and another provided an overview of the NAEP assessment. For the SAT, the purpose of the critical reading section was described, followed by a review of its features. The structure of the critical reading section was then outlined, including the timing of the test and the number of items. The types of reading passages and the nature of the text contained therein were also reviewed. The cognitive processes targeted through three different item types were then described. Finally, each panelist was given a copy of a sample SAT critical reading section from the 2007–2008 SAT Preparation Booklet. (This booklet was also sent to panelists in advance of the meeting.) A copy of the SAT presentation is included in Appendix D.

For the NAEP 12th-grade reading assessment, a description of the purpose of the test was given, along with a review of its features. The structure of the NAEP reading test, a survey assessment, was then described, and information on the number of items and the time given to each student was provided. The characteristics of passages found on the assessment, including length and type, were delineated, after which the types of cognitive targets were reviewed. Approaches used to assess meaning vocabulary were then described, and a summary of the types and distribution of items used on the test was given. A description was then provided of ways in which performance on the assessment was reported, both through average scores and through the achievement levels that are used to categorize groups of students. Lastly, each panelist was given a copy of the pre-publication edition of the Framework for the 2009 NAEP Reading Assessment (2007), as well as sample blocks from the 2008 NAEP 12th-grade reading pilot assessment. (The pre-publication framework was also sent to panelists in advance of the meeting.) A copy of the NAEP presentation is included in Appendix E.

Training in Content-Area Ratings

The panelists were then trained to make content-area comparability judgments. As described above, the first judgment was whether the SAT descriptor is covered by one or more NAEP descriptors, using a “Yes” or “No” rating. For those SAT descriptors for which the panelist responded “Yes,” he or she was asked to (a) indicate the strength of the NAEP-SAT comparability for that SAT descriptor, using a 3-point rating scale, ranging from (1) weak to (2) moderate to (3) strong, and to (b) identify the specific NAEP descriptors(s) that cover the SAT descriptor. A copy of the process training slides is contained in Appendix F.

Once the explanation of the process of making the content-area judgments was completed, the panelists were asked to read through both the SAT and NAEP descriptors. The panelists were then asked to provide ratings for the first three SAT critical reading descriptors. Panelists had the opportunity to discuss these initial judgments, helping them to come to a shared understanding of the judgment process. A copy of the content-area rating form is contained in Appendix G.

First Evaluation Form

Panelists were next given the Initial Evaluation/Ready-to-Proceed form. The panelists were asked to provide feedback on the extent to which they understood the purpose of the study, the degree to which the overview of the assessment was presented clearly, how clearly the steps to be followed in making ratings were presented, and the degree to which they understood what was expected of them in completing content-area ratings. They were also asked to sign off on whether they were ready to proceed to make the first set of content-area comparability judgments. A copy of the initial evaluation form is included in Appendix H.

Round 1 Ratings

After it was confirmed that all panelists felt ready to proceed, they were asked to complete the remaining round 1 ratings. These ratings were then entered into a spreadsheet by ETS staff, during which time the panelists took a break for lunch. Any ratings that were blank were given back to the panelists to complete before feedback and discussion began.

Feedback and Discussion

Selected portions of the spreadsheet were displayed to panelists to facilitate the discussion for those items for which there was the biggest split for Yes/No ratings. A sample feedback display is shown in table 6. Panelists were asked to describe reasons for their ratings, and they did so willingly. They were asked to make notes about ratings that they would like to change on the basis of the discussion.

Round 2 Ratings

Panelists were asked to record revised ratings based on the discussion and further examination of items not discussed. They were told that they were not required to revise any ratings if they did not wish to.

Overall Ratings

After round 2 ratings were completed, panelists were asked to complete their overall ratings. As described above, the ratings focused on the degree to which panelists believed that the tests covered the same type and the same range/breadth of reading skills, and whether a follow-up study comparing items on each test is justified. A summary of these ratings was reported to the panelists once ratings were completed and collected. A copy of the overall rating form is contained in Appendix I.

Final Evaluation

After turning in their overall ratings, panelists were given the final evaluation form. First, the panelists were asked whether the SAT and NAEP descriptors were sufficiently detailed to judge comparability. Next, they were asked several questions about the content-area comparability ratings—whether the rating form was easy to complete, whether the summary of ratings was presented clearly, whether the discussion of the summary of the ratings was informative, and whether the process of completing the

ratings was easy to follow. Finally, the panelists were asked a parallel set of questions about the overall ratings. A copy of the final evaluation form can be found in Appendix J.

Debriefing

A debriefing was held with panelists after the final survey was completed so that they could voice their opinions on different aspects of the process in an interactive manner.

Results

Content-Level Ratings

Summary-Level Results

In table 7, a summary of the ratings provided for each round, for each content area is presented. The total number and percentage of No, Yes (1), Yes (2), and Yes (3) ratings are given. For the Yes ratings, Yes (1) indicates weak comparability, Yes (2) indicates moderate comparability, and Yes (3) indicates strong comparability.

Figure 1 graphically presents the information for round 1. Content area B (Understanding Literary Elements) had the highest percentage of No ratings, at 36 percent, and content area C (Organization and Ideas) had the lowest, at 10 percent. For Yes (3) ratings, content area B (Understanding Literary Elements) had the highest percentage at 52 percent, and content area A had the lowest at 22 percent.

Data for round 2 ratings are graphed in Figure 2. The highest and lowest percentages of No ratings continued to be found in content areas B (decreasing to 32 percent) and C (increasing to 13 percent), respectively. For Yes (3) ratings, the highest percentages were found for content areas B (decreasing to 36 percent) and C (also at 36 percent), while the lowest percentage was again found for content area A (remaining at 22 percent).

The differences across rounds can be more easily seen in Figure 3. No changes were seen for content area A, and minimal changes were seen for content area C. Somewhat larger changes were seen for content areas D and E, but the greatest changes were seen for content area B. For content area B, No ratings decreased by 4 percent, Yes (3) ratings decreased by 16 percent, and Yes (2) ratings increased by 20 percent.

Panelist-Level Results

Data for round 1 are graphed by panelist in Figure 4. Panelist 5 had the largest percentage of No ratings with 51 percent; the panelists with the smallest percentage were Panelists 2 and 4 with 8 percent. The panelist with the highest percentage of Yes (3) ratings was Panelist 2 with 52 percent, while the panelist with the lowest percentage was Panelist 5 with 5 percent.

Round 2 data by panelist are graphed in Figure 5. Panelist 5 remains the panelist with the largest percentage of No ratings, though the percentage decreased from 51 percent to 44 percent. Panelist 4 continues to have the lowest percentage of No ratings, though it increased from 8 percent to 9 percent. Panelist 2 remains the panelist with the highest

percentage of Yes (3) ratings, which decreased from 52 percent to 45 percent. Panelist 5 continues to have the lowest percentage of Yes (3) ratings, remaining at 5 percent.

The differences across rounds at the panelist level can be more easily seen in figure 6. Panelists 2 and 3 showed the largest changes. Panelist 2 increased the percentage of No and Yes (2) ratings by 5 percent and 4 percent, respectively, and decreased the percentage of Yes (1) and Yes (3) ratings by 3 percent and 7 percent, respectively. Panelist 3 decreased the percentage of No ratings by 9 percent, while increasing the percentage of Yes (2) ratings by 9 percent.

Information about panelist changes is also given in table 8. Panelist 3 had the greatest number of changes with 11. Panelists 1 and 5 had the fewest with 5. By type of change, the greatest number were changed from No to Yes (1)³ and No to Yes (2), each with 9. There were 36 changes in all.

It is also of interest to examine the degree to which a descriptor's ratings were more likely to change if it had been discussed after round 1. After round 1 ratings were completed, entered, and summarized, ratings for 15 descriptors were discussed. As shown in table 9, three items from section A, four items from section B, one item from section C, three from section D, and four items from Section E were reviewed and had ratings related to them revised. One additional descriptor, A5, was reviewed, but no ratings were revised. There were also descriptors that had not been reviewed for which ratings were changed.

Overall Ratings

Summary-Level Results

Data on overall ratings are presented in table 10. All panelists agreed or strongly agreed that the content and breadth of the SAT and NAEP descriptors were comparable. All panelists indicated that there is sufficient overall overlap between the SAT and NAEP to justify conducting a more extensive alignment study at the item level.

Panelist-Level Results

Table 11 shows the overall ratings given by each of the panelists. Panelists 1 and 4 were the only ones to state that they strongly agreed that the two tests covered the same content. This did not appear to be linked to familiarity with a given test, since Panelist 1 was a former NAEP committee member, and Panelist 4 was an SAT committee member.

³ During discussion of round 1 results, panelists indicated they saw little distinction between a rating of No and a rating of Yes (1).

Panelist Feedback

Initial Evaluation

Panelist responses to the Initial Evaluation/Ready-to-Proceed form are shown in table 12. For each of the first four questions, three panelists strongly agreed and two panelists agreed that (a) they understood the purpose of the study, (b) the overview of the assessments was presented clearly, (c) the steps that they were to follow to make their content-area ratings were presented clearly, and (d) they understood what they would be expected to do to complete their content-area ratings. As indicated in table 13, Panelists 1 and 4 strongly agreed with all of the statements (they were also the two panelists who strongly agreed that the tests covered the same content), and Panelist 5 agreed with all of the statements.

In response to the last question, all panelists indicated that they were ready to proceed with round 1 ratings.

Final Evaluation

Panelist responses to the final evaluation form are shown in table 14. There were four sections to the form. In the first section, a question asked whether the SAT and NAEP skill statements were sufficiently detailed to judge comparability. Ratings were varied in response to this question—no panelists strongly agreed, three panelists agreed, two panelists disagreed, and no panelists strongly disagreed. As shown in table 15, the panelists who disagreed were Panelists 2 and 4, both of whom are SAT committee members.

In the second section, panelists were asked four questions in relation to the content-area comparability ratings. They were asked (a) whether the content-area rating form was easy to complete, (b) whether the summary of the content-area ratings was presented clearly, (c) whether the discussion of the summary of content-area ratings was informative, and (d) whether the process of completing the content-area ratings was easy to follow. All of the ratings were “strongly agree” except for one panelist each who answered “agree” to the form completion (a) and process (d) questions. Different panelists provided these ratings—Panelist 3 for form completion and Panelist 5 for process.

In the third section, the same questions asked in the previous section were repeated in relation to the overall ratings. All responses were “strongly agree” except for one panelist, Panelist 3, who responded “agree” for form completion.

In the final section, panelists were asked what those running the study should consider doing differently the next time this type of study is conducted. The responses are shown in table 16. Many of the comments related to the NAEP descriptors and their perceived lack of detail.

Discussion and Conclusions

Based on the results summarized above, the comparability study provided valuable information for the College Board and National Assessment Governing Board to consider when determining whether to proceed with an item-level alignment study. All panelists believed that such a study would be worthwhile given the level of comparability between the SAT performance characteristics descriptors and the NAEP preliminary achievement-level descriptions.

Responses to the evaluations conducted during the study indicate that the panelists understood the purpose of the study and the tasks they were to accomplish, and believed that different components of the study such as the rating form, discussion, and process for providing ratings were clear and helpful. The only aspect of the study about which the panelists did not feel as comfortable was the NAEP preliminary achievement-level descriptions, which they believed were not detailed enough. In addition, the lack of grade-level specific descriptions for the vocabulary subscale presented some challenges. However, these challenges did not prevent the panelists from making specific and overall ratings.

Panelists did make changes in response to feedback provided after round 1. However, it is possible that they may have made more changes had time allowed for more ratings of descriptors to be reviewed. Given the goal of the study, however, which was to provide a recommendation as to whether a more extensive item-level study should be conducted, it appears that the amount of time for ratings and review was sufficient.

In general, panelist ratings did not appear to be affected by their level of familiarity with a given test. Neither the responses on the evaluation forms nor the distribution of and changes in ratings seemed to be systematically related to whether a panelist was a NAEP or SAT committee member.

In summary, the study yielded panelist feedback that the content and breadth of the SAT and NAEP descriptors were comparable, and that there is sufficient overall overlap between the SAT and NAEP to justify conducting a more extensive alignment study at the item level.

References

- Bhola, D. S., Impara, J. C., & Buckendahl, C. W. (2003). Aligning tests with states' content standards: Methods and issues. *Educational Measurement: Issues and Practice*, 22(3), 21–29.
- College Board. (2007, June 11). *Performance category descriptions for the Critical Reading, Mathematics, and Writing sections of the Scholastic Assessment Test*. New York: Author.
- National Assessment Governing Board. (2007). *Reading framework for the 2009 National Assessment of Educational Progress: Pre-publication edition*. Washington, DC: Author.
- Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education*. Council of Chief State School Officers and National Institute for Science Education. Madison, WI: Wisconsin Center for Education Research, University of Wisconsin.
- Webb, N. M., Herman, J. L., & Webb, N. L. (2007). Alignment of mathematics and state-level standards and assessments: The role of reviewer agreement. *Educational Measurement: Issues and Practice*, 26(2), 17–29.

Table 1

Panelist Characteristics

Panelist Number	Committee	Occupational Setting	Region*
1	NAEP	K-12	West
2	SAT	K-12	Midwest
3	NAEP	Post-secondary	South
4	SAT	K-12	West
5	SAT	Post-secondary	South

*Region as defined by the U.S. Census.

Table 2

Comparison of Panelist and Committee Characteristics

Characteristic	Committee		
	Comparability Study Panel	NAEP Standing Committee	SAT Standing Committee
Occupational Setting			
K-12	3 (60%)	7 (47%)	4 (50%)
Post-Secondary	2 (40%)	8 (53%)	4 (50%)
Region*			
Northeast	0 (0%)	7 (47%)	1 (13%)
South	2 (40%)	2 (13%)	4 (50%)
Midwest	1 (20%)	5 (33%)	1 (13%)
West	2 (40%)	1 (7%)	2 (25%)

*Region as defined by the U.S. Census.

Table 3

*Number of Performance Characteristics Descriptors
for SAT Critical Reading Section*

Content Area	Number of Descriptors
A. Determining the Meaning of Words	20
B. Understanding Literary Elements	5
C. Organization and Ideas	14
D. Author's Craft	19
E. Reasoning and Inferencing	17
Total	75

Table 4

*Number of Preliminary Achievement Level Descriptions
for NAEP 12th Grade Reading Assessment*

Passage Type	Number of Descriptions
A. Literary	18
B. Informational	15
Total	33

Table 5

Meeting Agenda

8:00 a.m. – 8:30 a.m.	Continental Breakfast	
8:30 a.m. – 9:00 a.m.	Welcome Introduction Purpose of the Study	<i>Susan Loomis</i>
9:00 a.m. – 9:45 a.m.	Overview of the Assessments SAT Critical Reading Grade 12 NAEP Reading	<i>Barbara Elkins</i> <i>Nicole Beaulieu</i>
9:45 a.m. – 10:00 a.m.	Panel Ratings Procedures for Study	<i>Mary Pitoniak</i>
10:00 a.m. – Noon	Panelists Reviews	
Noon – 1:00 p.m.	Lunch	
1:00 p.m. – 2:00 p.m.	Summarize Ratings & Discuss Summarize Ratings Discussion	<i>Mary Pitoniak</i> <i>Panels</i>
2:00 p.m. – 2:30 p.m.	Revisit Ratings	
2:30 p.m. – 3:00 p.m.	Wrap-up and Discussion	
3:00 p.m.	Adjourn	

Table 6

Sample Display of Data from Round 1 Ratings

		Round 1 Ratings						
		# yes	# no	mean	SD	freq 1	freq 2	freq 3
Content Area	Descriptor							
A	1	5	0	2.0	1.0	2	1	2
A	2	5	0	1.8	0.8	2	2	1
A	3	4	1	1.3	0.5	3	1	0
A	4	5	0	2.4	0.9	1	1	3
A	5	2	3	1.0	0.0	2	0	0

Notes. In the columns showing number of Yes and No ratings, a cell is highlighted if it has a value of 3 in order to show ratings with the greatest split between Yes and No ratings. To calculate the mean ratings, “No” ratings were given a value of 0, the Yes (1) ratings a value of 1, the Yes (2) ratings a value of 2, and the Yes (3) ratings a value of 3.

Table 7

Ratings for Rounds 1 and 2

Content Area	Number of Descriptors	Number of Ratings*	Round 1				Round 2			
			No	Yes (1)	Yes (2)	Yes (3)	No	Yes (1)	Yes (2)	Yes (3)
A. Determining the Meaning of Words	20	100	31 (31%)	34 (34%)	13 (13%)	22 (22%)	31 (31%)	34 (34%)	13 (13%)	22 (22%)
B. Understanding Literary Elements	5	25	9 (36%)	1 (4%)	2 (8%)	13 (52%)	8 (32%)	1 (4%)	7 (28%)	9 (36%)
C. Organization and Ideas	14	70	7 (10%)	18 (26%)	19 (27%)	26 (37%)	9 (13%)	17 (24%)	19 (27%)	25 (36%)
D. Author's Craft	19	95	20 (21%)	17 (18%)	30 (32%)	28 (29%)	14 (15%)	16 (17%)	35 (37%)	29 (31%)
E. Reasoning and Inferencing	17	85	23 (27%)	12 (14%)	28 (33%)	22 (26%)	17 (20%)	19 (22%)	26 (31%)	20 (24%)
Total	75	375	90 (24%)	82 (22%)	92 (25%)	111 (30%)	79 (21%)	87 (23%)	100 (27%)	105 (28%)

*There were 5 ratings (one for each panelist) for each descriptor. The numbers and percentages shown for each round are based on ratings, not descriptors.

Note. Yes (1) indicates weak comparability, Yes (2) indicates moderate comparability, and Yes (3) indicates strong comparability.

Table 8

Number of Changes in Ratings from Round 1 to Round 2 by Panelist

		Panelist					
Direction of Change		1	2	3	4	5	Total
No to Yes	No to Yes (1)	1		2	1	5	9
	No to Yes (2)	2		7			9
	No to Yes (3)	1					1
Yes to No	Yes (1) to No		2	2			4
	Yes (2) to No	1	1		1		3
	Yes (3) to No		1		1		2
Decrease in level of Yes rating	Yes (2) to Yes (1)				3		3
	Yes (3) to Yes (2)		4		1		5
Total		5	8	11	7	5	36

Note. The total number of ratings made in each round was 375.

Table 9

Descriptors Reviewed and Ratings Revised from Round 1 to Round 2

Content Area	Total Number of Descriptors	Number of Descriptors Reviewed	IDs of Descriptors Reviewed	Number of Descriptors with Revised Ratings	IDs of Descriptors Revised*
A. Determining the Meaning of Words	20	3	A5, A12, A16	2	A12, A16
B. Understanding Literary Elements	5	4	B2, B3, B4, B5	4	B2, B3, B4, B5
C. Organization and Ideas	14	1	C2	1	C2
D. Author's Craft	19	3	D1, D2, D10	5	D1, D2, D10, D18, D19
E. Reasoning and Inferencing	17	4	E5, E8, E13, E14	8	E3, E5, E8, E13, E14, E15, E16, E17
Total	75	15		20	

*IDs in **bold** are for descriptors that were both reviewed and revised.

Table 10

Summary of Overall Ratings

Topic	Rating			
	Strongly Agree	Agree	Disagree	Strongly Disagree
Content	2	3	0	0
Breadth	0	5	0	0
	Yes	No		
Future Study	5	0		

Table 11

Overall Ratings by Panelist

Topic	Panelist				
	1	2	3	4	5
Content	strongly agree	agree	agree	strongly agree	agree
Breadth	agree	agree	agree	agree	agree
Future Study	yes	yes	yes	yes	yes

Table 12

Summary of Responses to Initial Evaluation Form

Topic	Rating			
	Strongly Agree	Agree	Disagree	Strongly Disagree
I understand the purpose of the study.	3	2	0	0
The overview of the assessments was presented clearly.	3	2	0	0
The steps that I am to follow to make my content-area ratings were presented clearly.	3	2	0	0
I understand what I will be expected to do to complete my content-area ratings.	3	2	0	0
	Yes		No	
I am ready to proceed and to make my first set of content-area comparability judgments.	5	0		

Table 13

Responses to Initial Evaluation Form by Panelist

Topic	Panelist				
	1	2	3	4	5
Understand purpose	strongly agree	agree	strongly agree	strongly agree	agree
Overview presented clearly	strongly agree	agree	strongly agree	strongly agree	agree
Steps presented clearly	strongly agree	strongly agree	agree	strongly agree	agree
Understand how to complete ratings	strongly agree	agree	strongly agree	strongly agree	agree
Ready to Proceed	yes	yes	yes	yes	yes

Table 14

Summary of Responses to Final Evaluation Form

Statement	Rating				
	Strongly Agree	Agree	Disagree	Strongly Disagree	No Response
The SAT and NAEP skill statements were sufficiently detailed to judge comparability.	0	3	2	0	0
<i>Content-Area Comparability Ratings</i>					
The content-area rating form was easy to complete.	3	1	0	0	1
The summary of our content-area ratings was presented clearly.	5	0	0	0	0
The discussion of the summary of content-area ratings was informative.	5	0	0	0	0
The process of completing the content-area ratings was easy to follow.	4	1	0	0	0
<i>Overall Ratings</i>					
The overall rating form was easy to complete.	4	1	0	0	0
The summary of our overall ratings was presented clearly.	5	0	0	0	0
The discussion of the summary of overall ratings was informative.	5	0	0	0	0
The process of completing the overall ratings was easy to follow.	5	0	0	0	0

Table 15

Responses to Final Evaluation Form by Panelist

Area	Topic	Panelist				
		1	2	3	4	5
General	Descriptors detailed	agree	disagree	agree	disagree	agree
Content-Area Ratings	Form completion	strongly agree	strongly agree	agree	strongly agree	--
	Ratings summary	strongly agree	strongly agree	strongly agree	strongly agree	strongly agree
	Discussion informative	strongly agree	strongly agree	strongly agree	strongly agree	strongly agree
	Process easy to follow	strongly agree	strongly agree	strongly agree	strongly agree	agree
Overall Ratings	Form completion	strongly agree	strongly agree	agree	strongly agree	strongly agree
	Ratings summary	strongly agree	strongly agree	strongly agree	strongly agree	strongly agree
	Discussion informative	strongly agree	strongly agree	strongly agree	strongly agree	strongly agree
	Process easy to follow	strongly agree	strongly agree	strongly agree	strongly agree	strongly agree

Table 16

Responses to Open-Ended Question on Final Evaluation Form by Panelist

Panelist	Response
1	The NAEP descriptors need to be fleshed out and developed to describe more precisely the skills and content described in the NAEP framework.
2	Descriptor rating: NAEP's lack of detail caused many disagreements. (1) Ask "what should we" instead of "what we should." (2) You may want to make more time for discussing standard deviations (green boxes). Otherwise, I thought the meeting was well-managed.
3	The practice items were vocabulary. I think it would have been good to do 1 or 2 practice items from each of the categories (A-E). Obviously vocabulary descriptors need to be developed.
4	The only part that was difficult had to do with the language of the NAEP descriptors. Some of the verbs seemed too similar to make a distinction (i.e., "describe the character's motivation" vs "analyze character's motivation"), while at other times the descriptor seemed far too broad (i.e., B15). I think that once the NAEP language is more specific, then the comparability study will be more effective.
5	Consider selecting, consolidating, and streamlining the NAEP material needed only for comparability work—printed material in particular. I found the oral and PowerPoint presentation of essential background information was sufficient.

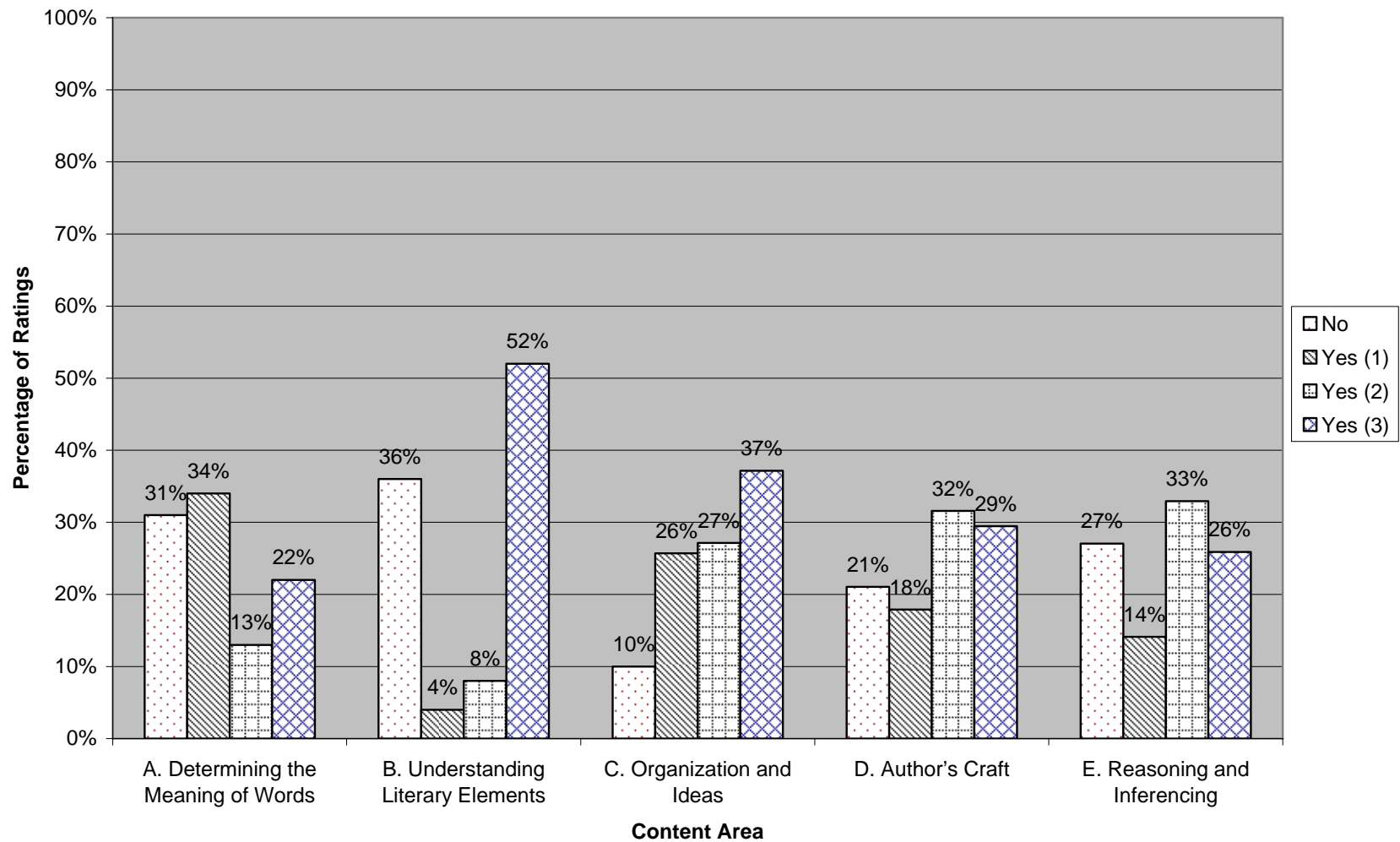


Figure 1. Round 1 Ratings by Content Area.

For each content area, the percentage of ratings falling into each of the four rating categories is shown.

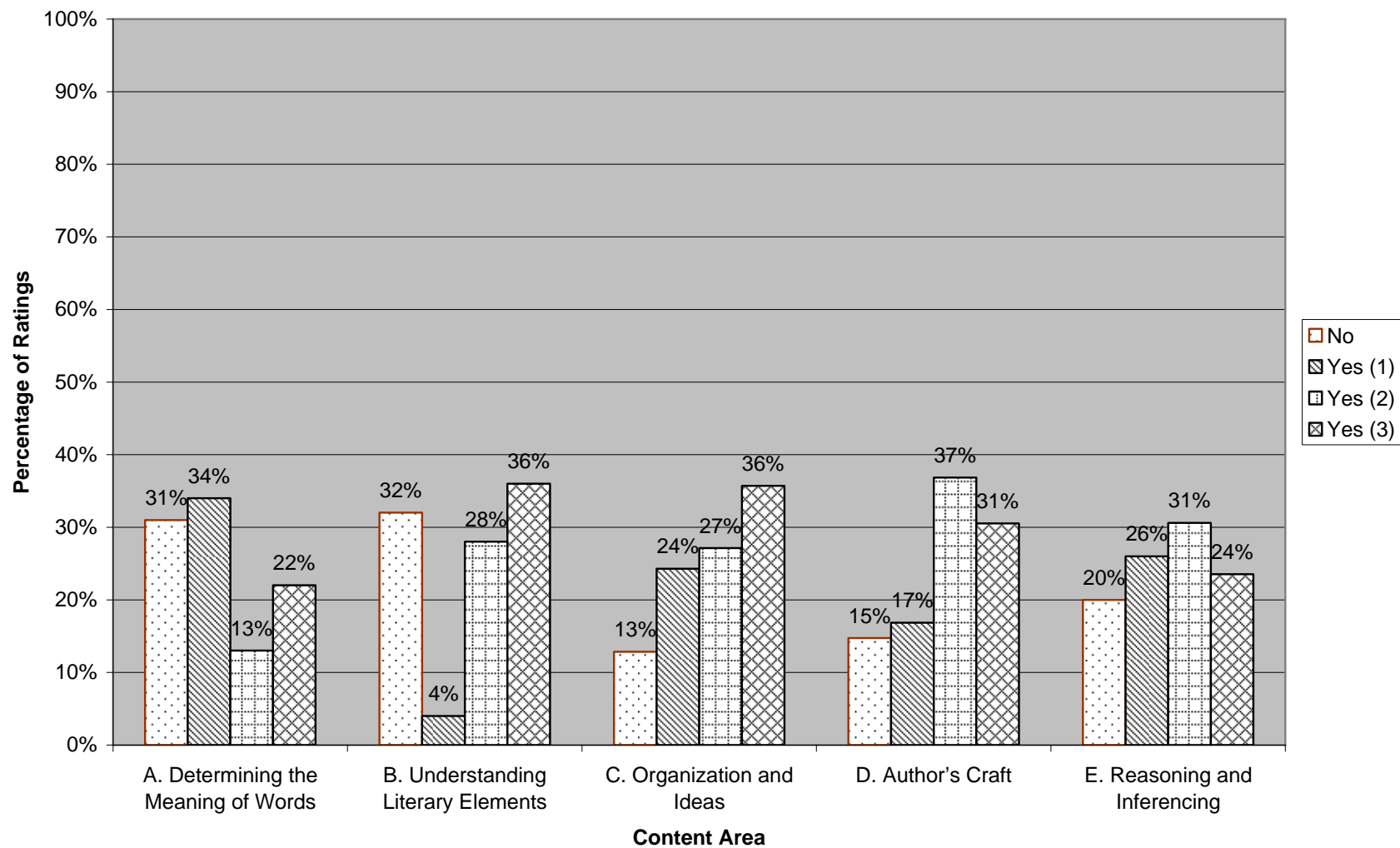


Figure 2. Round 2 Ratings by Content Area.

For each content area, the percentage of ratings falling into each of the four rating categories is shown.

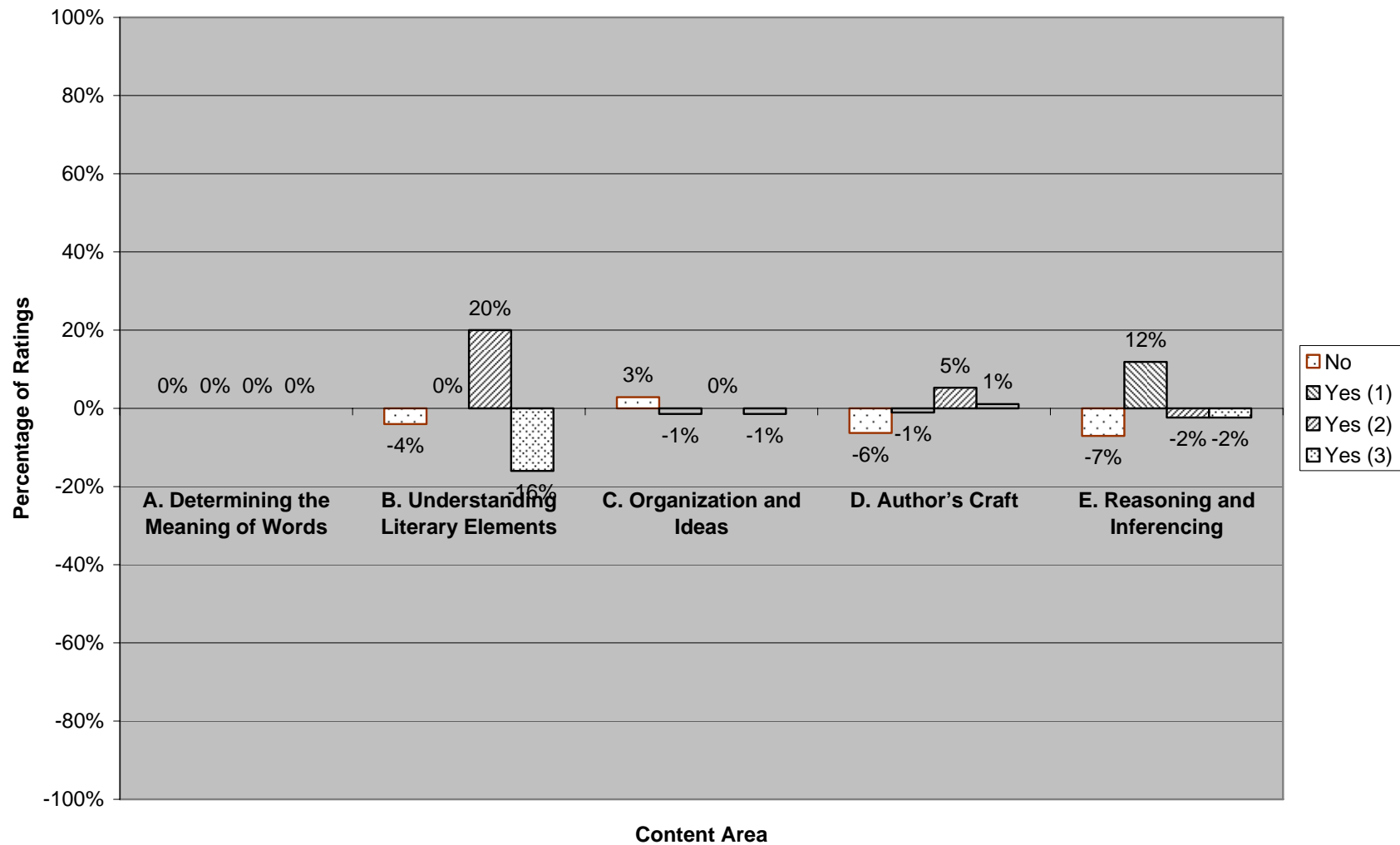


Figure 3. Changes in Ratings Over Rounds by Content Area.
 For each content area, the change in percentage of ratings falling into each of the four rating categories is shown.

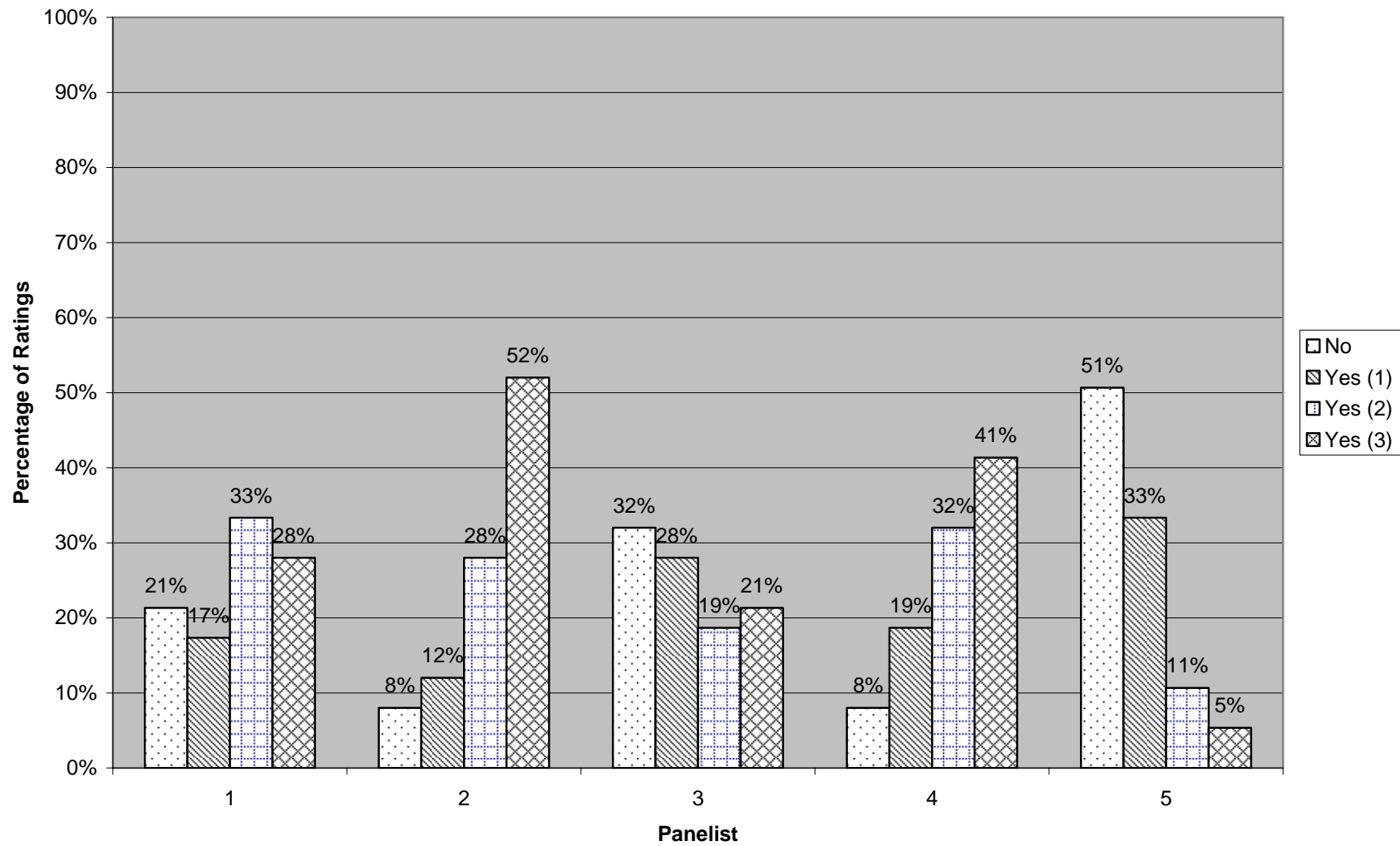


Figure 4. Round 1 Ratings by Panelist.
 For each panelist, the percentage of ratings falling into each of the four rating categories is shown.

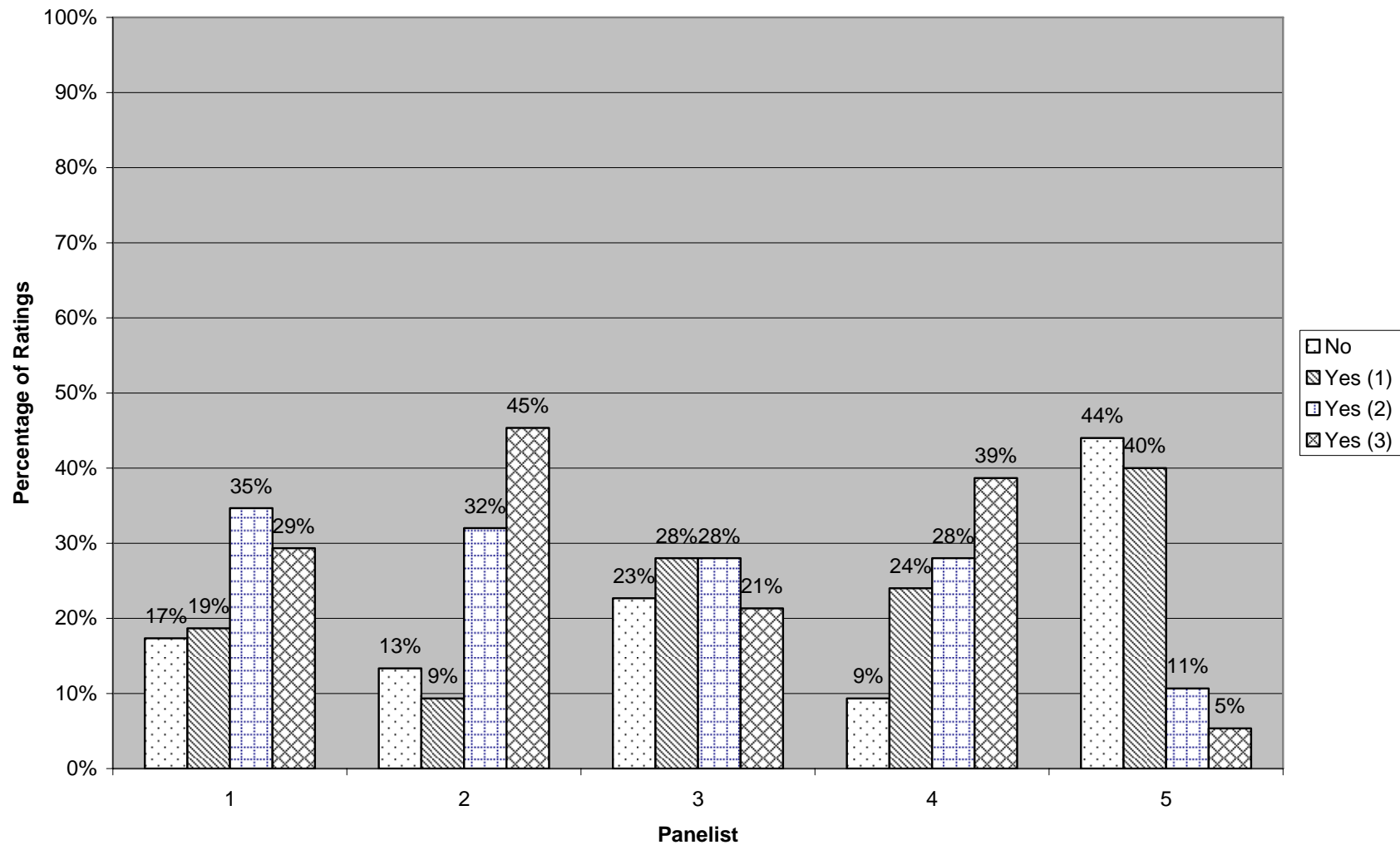


Figure 5. Round 2 Ratings by Panelist.
 For each panelist, the percentage of ratings falling into each of the four rating categories is shown.

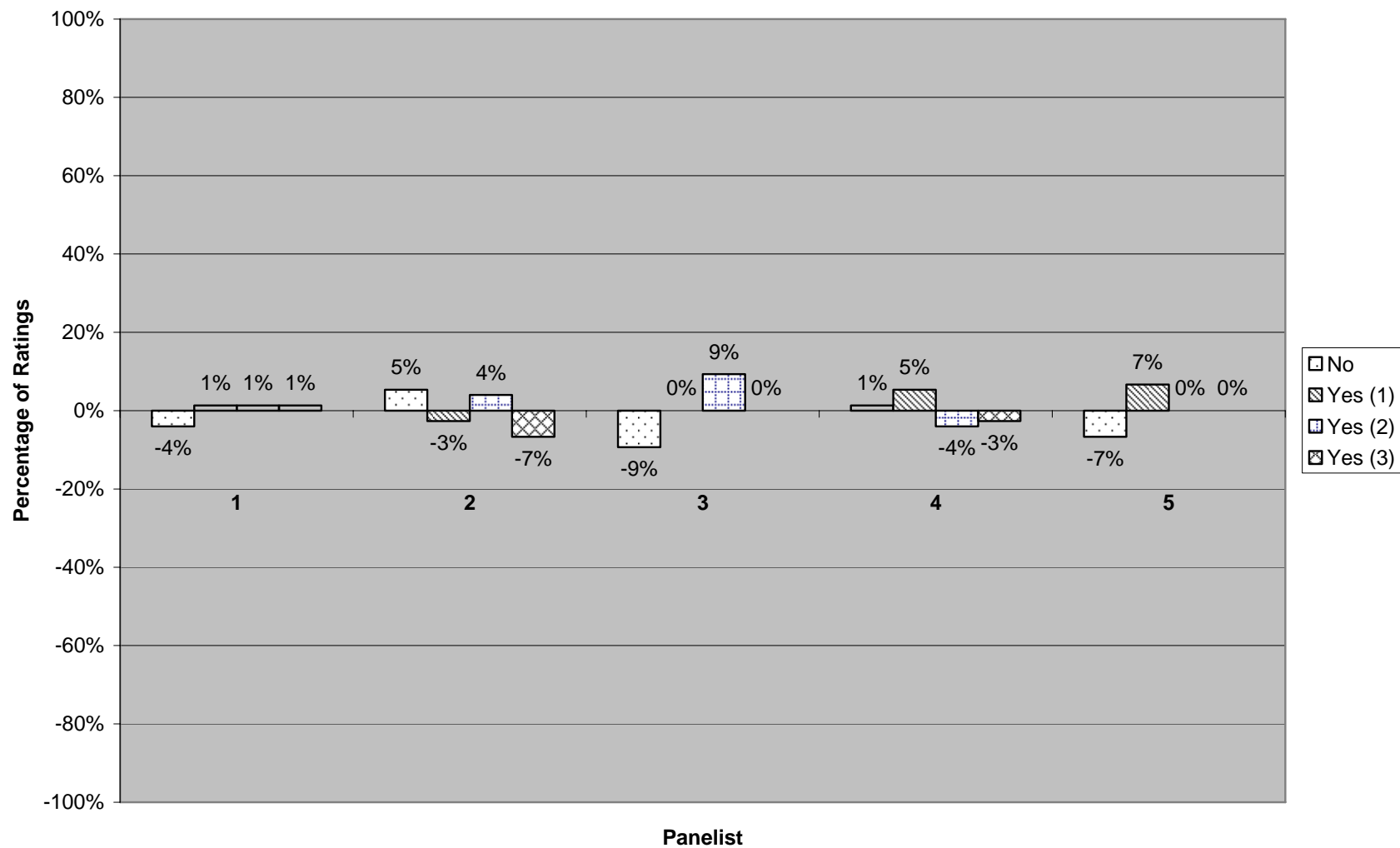


Figure 6. Changes in Ratings Over Rounds by Panelist.
 For each panelist, the change in percentage of ratings falling into each of the four rating categories is shown.

Appendix A

SAT Critical Reading Performance Category Descriptors

Performance Characteristics Descriptors

Critical Reading

A. Determining the Meaning of Words

1. Determine the meaning of words in a simple sentence by using context clues including familiar phrases and other vocabulary in the sentence
2. Use context clues when selecting missing vocabulary at the sentence level
3. Use knowledge of root words, prefixes, and suffixes when selecting missing vocabulary at the sentence level
4. Use the context of a sentence or a short section of text to clarify the meaning of unknown words or to select the appropriate meaning of familiar and simple words that have multiple meanings
5. Use knowledge of root words to determine the meaning of words needed to complete a compound or complex sentence
6. Recognize and understand less common words and specialized vocabulary (terms used in a particular occupation or field of study)
7. Use context clues (such as an embedded definition) to select missing vocabulary at the sentence level
8. Use the context of a sentence or a short section of text to clarify the meaning of unknown words (when definitions may or may not be embedded in the text) or to select the appropriate meaning of familiar and simple words that have multiple meanings)
9. Use sentence structure to negotiate the meaning of the sentence
10. Make sense of complex sentences with logical constructions that include terms such as *but*, *although*, *or*, *if*, *then*, and *not*
11. Use the context of a sentence or larger section of text to determine the meaning of unknown words or to differentiate among multiple possible meanings of words
12. Understand how syntax (the arrangement of words and phrases in a sentence) influences the relationship among words and ideas within a sentence
13. Demonstrate increased comprehension of specialized vocabulary
14. Understand familiar words in unfamiliar contexts and differentiate among multiple possible meanings for words in unfamiliar contexts
15. Understand sophisticated and specialized vocabulary
16. Determine the meaning of a word when there is little or no supporting context
17. Negotiate complex syntax (the arrangement of words and phrases in a sentence), and integrate ideas within and across sentences
18. Understand how words can sometimes be used in unusual ways that directly refute common usage
19. Access broad and extensive vocabulary within complex syntactical structures and in a variety of contexts
20. Analyze the context of a sentence or larger sections of text to clarify the meaning of unknown words, differentiate among multiple possible meanings of words, detect nuances, and infer connotations

Performance Characteristics Descriptors

Critical Reading

B. Understanding Literary Elements

1. Identify nuances and attitudes of characters
2. Determine characterization from dialogue, thoughts and actions, interactions among characters, and narrative perspective
3. Analyze characters' function in a narrative
4. Interpret dialogue from a character's or from the narrator's perspective
5. Analyze the roles and relationships among characters and between character and narrator

Performance Characteristics Descriptors Critical Reading

C. Organization and Ideas

1. Understand the central idea(s) in a simple text or in a short section of a longer text
2. Determine the main idea of a text and apply it to a different context
3. Understand the relationship of ideas within and across different texts
4. Integrate information from short sections of different texts
5. Integrate ideas within and across sentences and texts
6. Comprehend generalizations about texts
7. Analyze and compare concepts across texts
8. Draw text-based conclusions beyond the main idea
9. Determine the function of a selected portion of text within a longer text
10. Analyze main ideas and concepts within and across complex and sometimes opposing texts
11. Compare and contrast explicit and implicit supporting ideas across texts
12. Recognize components of an author's argument within a text
13. Analyze context, sentence structure, and sentence variation to construct meaning within and across sentences and texts
14. Interpret multiple layers of a text

Performance Characteristics Descriptors

Critical Reading

D. Author's Craft

1. Use vocabulary clues to determine the tone of a text (the author's attitude toward the subject of the text and toward the audience)
2. Recognize elements of figurative language (such as metaphor) in a text
3. Identify an author's purpose for writing
4. Identify and describe the effects of literary devices used to achieve a specific purpose
5. Infer the author's opinion concerning the central ideas in a text
6. Use tone to infer an author's unstated assumptions
7. Use context clues to identify an author's rhetorical purpose (for example, to persuade the audience) in a short section of text
8. Infer an author's purpose for writing
9. Recognize the use of irony and the effects of other sophisticated literary devices, such as symbolism, in a text
10. Determine the function of words and devices in limited sections of text
11. Analyze how an author achieves specific effects using rhetorical devices and strategies
12. Analyze an author's explicit and implicit purposes for writing
13. Analyze the effects of an author's rhetorical and stylistic choices
14. Distinguish among opinion, fact, conjecture, and hypothesis in a text
15. Recognize how an author uses evidence to support a particular position
16. Recognize subtleties and differences in tone, such as the use of humor or irony to achieve a specific effect
17. Analyze the overall purpose of an author's text
18. Analyze how sophisticated rhetorical devices support an author's purposes
19. Analyze how the sophisticated use of literary devices and figurative language (such as extended metaphors, complicated analogies, and symbolic images) achieve specific effects

Performance Characteristics Descriptors

Critical Reading

E. Reasoning and Inferencing

1. Perform clear, simple steps of reasoning
2. Recognize a general idea, such as a paraphrase, that is supported by separate but related points in different sentences
3. Perform multiple steps of reasoning
4. Make multiple, layered inferences
5. Make layered inferences and apply those inferences to different but related situations
6. Make a connection between one part of a text and a later part of that same text to enhance comprehension
7. Perform complex reasoning tasks on short sections of text
8. Determine an author's unstated assumptions and develop inferences from explicit evidence in different sections of a text
9. Draw multiple extended inferences that require several steps of reasoning
10. Draw inferences based on implications throughout a text
11. Consider the entire text when making inferences, linking information to ideas both before and after a specific section
12. Integrate both general and detailed information across texts
13. Make inferences when there is no explicit reinforcement in the text or when information is missing, and use those inferences to draw further conclusions about the text
14. Apply conclusions drawn from a text to other contexts, understanding similar or analogous situations in the process
15. Identify an author's unstated assumptions and draw further conclusions about the text based on these assumptions
16. Analyze and relate multiple perspectives on similar topics across texts
17. Compare and contrast deeply embedded details or ideas across texts

Appendix B

NAEP 12th Grade Reading Preliminary Achievement Level Descriptions

Preliminary Achievement Level Descriptors

Grade 12 NAEP Reading

A. Literary

1. Interpret textually explicit information
2. Make inferences
3. Make inferences that describe problem and solution, cause and effect
4. Make complex inferences
5. Describe character's motivation
6. Analyze character motivation
7. Evaluate character motivation
8. Recognize alternative interpretations or point of view
9. Critique point of view
10. Explain the theme
11. Integrate ideas to determine theme
12. Explain how the message is affected by the genre
13. Explain thematic connections across literary texts
14. Examine relationships between theme, setting, or character
15. Identify elements of an author's style
16. Interpret mood or tone
17. Analyze how an author uses literary devices to convey meaning
18. Analyze and evaluate how an author uses literary devices to convey meaning

B. Informational

1. Summarize the main idea
2. Identify author's purpose
3. Identify key details
4. Find evidence in support of an argument
5. Integrate information from a variety of sources
6. Identify causal relations
7. Analyze causal relations
8. Draw conclusions
9. Determine unstated assumptions
10. Analyze point of view
11. Critique point of view
12. Judge the logic, coherence, or credibility of an argument
13. Evaluate the quality of supporting evidence
14. Critique the presentation of information
15. Evaluate the quality of counterarguments within and across texts

Appendix C

National Assessment Governing Board PowerPoint Presentation on Preparedness



Reporting Preparedness for Grade 12 NAEP

**Goals of the Governing Board for
Reading and Mathematics NAEP**

**Susan Cooper Loomis
Assistant Director, Psychometrics**



Overview

- Recommendation to the Board in 2004 by a national commission appointed to address issues related to 12th grade NAEP
- Board commissioned papers, appointed Ad Hoc Committees, and a Technical Panel on 12th Grade Preparedness Research to help address the goal of reporting preparedness for 12th grade NAEP



Members of the Technical Panel on 12th Grade Preparedness Research

Michael Kirst (Chair)

Stanford University

John Campbell

University of Minnesota

David T. Conley

University of Oregon

Michael Kane

**National Conference of
Bar Examiners**

Mark David Milliron

Catalyze Learning International

Robert Mislevy

University of Maryland

George C. Thornton, III

Colorado State University



Preparedness for Post-Secondary Activities

- ❑ Higher Education and Workplace (job or job training programs—civilian or military)
- ❑ Academic preparation, not behaviors that are known to be important indicators of readiness for college or workplace
- ❑ Preparedness means “remediation free;” eligible for placement in college credit-bearing course or job/job training program in reading/mathematics



Types of Prospective Studies

- Content alignment between NAEP and other assessments: a necessary first step
- Identifying NAEP scores and score ranges indicating preparedness via:
 - Judgments by subject matter experts
 - Statistical relationships with performance on other assessments



Content Alignment Studies

- ❑ Evaluate extent of content overlap between NAEP and other assessments
- ❑ Small-scale studies as a preliminary step
- ❑ Full-scale studies guide later statistical analyses



Purpose of this Study

- Get an early signal on the feasibility of using the SAT as an indicator of preparedness for NAEP
 - Want to develop statistical relationship to report SAT scores associated with preparedness in reading for placement in college credit courses (English, social sciences) and preparedness for workplace training program
 - Want to report percentage of students on NAEP that score at a level indicative of College Success, based on SAT data



Materials for Preliminary Alignment Study

- Framework for NAEP Reading
- Preliminary Achievement Levels Descriptions
 - Developed as detailed lists of reading skills and abilities students *should* have to demonstrate performance at each level of achievement.
 - *Basic = partial mastery*
 - *Proficient = solid academic performance; competency over challenging subject matter*
 - *Advanced = superior performance*



Purpose of Preliminary Achievement Levels Descriptions

- ❑ Developed by Framework Development Panel
- ❑ Developed to inform passage selection and item development to help assure coverage of appropriate skill levels for each grade
- ❑ Provide a “check list” for use in development of narrative descriptions that are used in setting achievement levels and reporting NAEP results



Questions?

Contact Information:

Susan.Loomis@ed.gov

(202) 357-6940

Appendix D

SAT Critical Reading Content Overview PowerPoint Presentation



Overview of SAT Critical Reading

Barbara Elkins, ETS
January 28, 2008



What does the SAT do?

- Provides an objective measurement of students' readiness for college
- Helps colleges make fair and informed decisions about applicants

The SAT is a proven, reliable indicator of college success



Overview of the SAT Critical Reading Assessment

- Reading comprehension assessed through passages of different lengths from different genres
- Reading comprehension and vocabulary knowledge assessed through sentence completion items
- Cognitive processes targeted through different types of sentence completion, extended reasoning, and literal comprehension items
- Test comprised of five-option multiple choice items



What is SAT Critical Reading like for students?

- Three sections of a ten-section operational test
- 67 reading items: 48 reading passage items, 19 sentence completions
- 70 minutes allotted for reading sections; 3 hours and 45 minutes allotted for all ten sections
- Most students take SAT reasoning test in both junior and senior years of high school



SAT Reading Passages

- Authentic text students are likely to encounter in college courses
- Passages are 100-850 words
- Each operational test includes the following:
 - three long passages (500-, 650-, and 800-words) and two paired passages (one 200-word pair and one long pair)
 - two 100-word paragraphs



Types of Text

Each test includes:

- One fiction passage, usually from 20th or 21st century
- One passage from each content area:
 - **Humanities** (art, literature, philosophy, music, architecture, folklore, drama, etc.)
 - **Social Science** (history, economics, business, media, government, culture, sociology, etc.)
 - **Natural Science** (biology, chemistry, astronomy, physics, medicine, geology, agriculture, ecology, etc.)
- Expository passages are either analytical or persuasive



Cognitive Processes Targeted through Different Item Types

- **Sentence Completions**
 - Regular: assess logical reasoning and vocabulary knowledge
 - Definitional: assess vocabulary knowledge



Cognitive Processes Targeted through Different Item Types

- Extended Reasoning Items:

- Primary Purpose
- Rhetorical Strategies
- Implication and Evaluation
- Tone and Attitude
- Application and Analogy

* Extended reasoning items, which test different kinds of reasoning skills, are used in both expository and fiction passages



Cognitive Processes Targeted through Different Item Types

- Literal Comprehension items
 - Assess understanding of explicitly stated points crucial to the logical or rhetorical development of a passage

- Vocabulary in Context items
 - Assess understanding of a word's meaning within the context of a sentence or reading passage



Sample SAT Critical Reading Test

- from 2007-2008 SAT Preparation Booklet



Questions?
Comments?

Appendix E

NAEP 12th Grade Reading Content Overview PowerPoint Presentation



Overview of the NAEP Reading Assessment at Grade 12

Nicole Beaulieu, ETS
January 28, 2008



What is NAEP?

- Nation's Report Card
- Only nationally representative sample and continuing assessment of what America's students know and can do
- Survey based on representative sample; no individual scores reported



Overview of the NAEP Reading Assessment

- Administered at grades 4, 8, and 12
- Assesses students' reading comprehension of two types of text: Literary and Informational
- Comprehension processes defined by cognitive targets
- Assesses meaning vocabulary systematically
- Combination of multiple-choice and constructed-response questions



What is NAEP Reading like for students?

- Long test, short booklet for students
- Students take two 25-minute blocks
 - (A “block” is a reading passage or pair of passages and the accompanying 10-12 items)
- Full grade 12 reading assessment contains 13 comprehension blocks and 4 vocabulary half-blocks
- Some of those blocks are cross-grade (administered at grades 8 and 12)



Passages on NAEP Reading

- Authentic text students are likely to encounter in school and in out-of-school situations
- Passages are 500-1500 words
- 20-30% of blocks are based on paired passages



Types of Text at Grade 12

- Literary (30%)
 - Fiction (20%)
 - Literary Nonfiction (5%)
 - Poetry (5%)

- Informational (70%)
 - Exposition (30%)
 - Argumentation/Persuasive Text (30%)
 - Procedural Text and Documents (10%)

FW p. 15-29



Cognitive Targets

- Locate/Recall (20%)
- Integrate/Interpret (45%)
- Critique/Evaluate (35%)

FW p. 36-41



Locate/Recall (20%)

- Identify textually explicit information, such as:
 - Facts
 - Supporting details
 - Setting
 - Sequence of events
 - Topic sentence
 - Causal relations

(FW p. 37 & 40)



Integrate/Interpret (45%)

- Make complex inferences to:
 - Describe problem and solution
 - Compare or connect ideas
 - Determine unstated assumptions in an argument
 - Examine relations between theme and setting
 - Interpret characters' motivation and decisions
 - Draw conclusions and provide supporting information

(FW p. 38 & 40)



Critique/Evaluate (35%)

- Consider text critically to:
 - Judge author's craft and technique
 - Evaluate the author's perspective or point of view
 - Analyze the presentation of information
 - Evaluate the author's selection of language to influence readers
 - Determine the quality of counterarguments within and across texts

(FW pp. 38-40)



Vocabulary

- Goal is to measure students' **meaning** vocabulary:
The application of one's understanding of word meanings to passage comprehension
- Words are presented in context of passage, not in isolation
 - Characterize vocabulary of mature language users at that grade level
 - Necessary for comprehension of local context; linked to central ideas
 - Not technical (i.e. photosynthesis)
- Two vocabulary items embedded in each comprehension block
- Separate vocabulary blocks spiraled with comprehension blocks

FW pp. 33-36



Item Types

Distribution of assessment time at grade 12 by item type:

- Multiple Choice (40%)
- Constructed Response
 - Short Constructed Response (45%)
 - Extended Constructed Response (15%)



NAEP Achievement Levels

- NAEP results are reported in two ways:
 - in terms of average scores for groups of students
 - as percentages of students who attain each of the three achievement levels: Basic, Proficient, and Advanced
- Preliminary achievement level descriptors describe the important reading skills that students at grade 12 should have mastered at the Basic, Proficient, and Advanced levels
- Preliminary achievement level descriptors are defined separately for Literary and Informational text

FW p. 47



Sample Grade 12 Pilot Blocks (Handout)

- One Literary/Fiction block
- One Informational/Exposition & Procedural block
(paired passages)
- Two vocabulary half-blocks



Questions?
Comments?

Appendix F

Rating Process Training PowerPoint Presentation



NAEP Grade 12 Reading and SAT Critical Reading Assessments: Are They Comparable?

**Mary J. Pitoniak, ETS
January 28, 2008**



What Do We Mean By Comparability?

- It refers to whether the content covered by the SAT Critical Reading test is also covered by the NAEP grade 12 Reading test.
- So, for example, if SAT addresses a student's ability to *infer author's intended purpose*, the question is: Does the NAEP test also address this ability?
- Comparability is based on your expert judgment:
 - Do you believe that what's on the SAT Critical Reading test is also on the NAEP grade 12 Reading test?



What Do We Mean By Comparability? *(continued)*

- The question of comparability goes in one direction:
Does NAEP cover what's on the SAT?
- It does **not** ask:
Does SAT cover what's on the NAEP?



What Materials Will We Use To Judge Comparability?

- For SAT, statements of the reading skills needed to answer the range of questions on the test—termed *performance characteristics descriptors*
- Descriptors are organized by 5 main categories:
 - **Determining the Meaning of Words**
 - **Understanding Literary Elements**
 - **Organization and Ideas**
 - **Author’s Craft**
 - **Reasoning and Inferencing**



What Materials Will We Use To Judge Comparability? *(continued)*

- For NAEP, statements of the reading skills that students at grade 12 should have—termed *preliminary achievement level descriptors*
- *Descriptors are organized by the two text types, literary and informational*
- *Descriptors do not address vocabulary skills*



What Are The Steps To Follow For Judging Comparability?

- First, take 15 minutes to review the SAT and NAEP descriptors
- Take out your judgment form
- Start with the first SAT descriptor
- Then refer to the NAEP preliminary achievement level descriptors to determine if the SAT descriptor is covered by NAEP
 - Yes or No



What Are The Steps To Follow For Judging Comparability? *(continued)*

- If you responded “yes,” then judge the strength of the NAEP-SAT comparability for that SAT descriptor:
 - **Weak, Moderate, or Strong**
- If you responded “yes,” then identify the specific NAEP descriptor(s) that cover the SAT descriptor
 - **Write down the NAEP descriptor number(s)**



What Are The Steps To Follow For Judging Comparability? *(continued)*

- We will collect your judgments and summarize them
- We will then share the summary and ask you to discuss the results and to share your perspectives
 - For which descriptors does there seem to be more or less convergence of judgment?
 - What led you to see or not see comparability?



What Are The Steps To Follow For Judging Comparability? *(continued)*

- After discussion, consider if you want to revise one or more of your judgments
 - You are not required to change your judgments, but this is your opportunity to do so.



What Are The Steps To Follow For Judging Comparability? *(continued)*

- After the second round of judgments, you will be asked to respond to 3 overall questions:
 - Whether the two tests cover the same types of reading skills
 - 1 (Strongly Agree) to 4 (Strongly Disagree)
 - Whether the two tests cover the same range/breadth of reading skills
 - 1 (Strongly Agree) to 4 (Strongly Disagree)
 - Based on your evaluation of comparability, do you think that a follow-up study comparing items on each test is justified?
 - Yes or No



What Are The Steps To Follow For Judging Comparability? *(continued)*

- Complete a final evaluation form regarding your experience with the process.



Before You Begin Your Judgments

- Any questions?
- Complete the first evaluation form and I will collect it.

Appendix G
Content-Area Rating Form

Panelist #: _____

NAEP/SAT Comparability Study Reading

Panelist Rating Form—Content-Area Ratings

On this form you will provide two types of ratings.

- The first rating is whether the content described by the SAT Critical Reading performance characteristics descriptor is covered by preliminary achievement level descriptors in the NAEP Reading framework. This is a yes or no rating.
- The second rating should be made only for those SAT descriptors that you have rated as a “yes,” that they are covered by preliminary achievement level descriptor(s) in the NAEP framework. This is a rating of the extent to which the NAEP preliminary achievement level descriptor(s) cover the same range of content/topics as the SAT descriptors, and is on a 3-point scale. A rating of “1” indicates weak alignment, “2” indicates moderate alignment, and “3” indicates strong alignment.
- A column is also provided in which to indicate which NAEP preliminary achievement level descriptor(s) cover the same content as the SAT descriptor. For example, an indication of *A5* would indicate that this NAEP preliminary achievement level descriptor — *Describe character’s motivation*—covers the same content as the given SAT descriptor.

Panelist Rating Form—Content-Area Ratings Reading

SAT Reading Descriptor		NAEP Objective(s)	Round 1 Ratings			Round 2 Ratings						
			Overall Coverage	Level of Comparability*		Overall Coverage	Level of Comparability*					
A <i>Determining the Meaning of Words</i>												
1	Determine the meaning of words in a simple sentence by using context clues including familiar phrases and other vocabulary in the sentence		Yes	No	1	2	3	Yes	No	1	2	3
2	Use context clues when selecting missing vocabulary at the sentence level		Yes	No	1	2	3	Yes	No	1	2	3
3	Use knowledge of root words, prefixes, and suffixes when selecting missing vocabulary at the sentence level		Yes	No	1	2	3	Yes	No	1	2	3
4	Use the context of a sentence or a short section of text to clarify the meaning of unknown words or to select the appropriate meaning of familiar and simple words that have multiple meanings		Yes	No	1	2	3	Yes	No	1	2	3
5	Use knowledge of root words to determine the meaning of words needed to complete a compound or complex sentence		Yes	No	1	2	3	Yes	No	1	2	3
6	Recognize and understand less common words and specialized vocabulary (terms used in a particular occupation or field of study)		Yes	No	1	2	3	Yes	No	1	2	3

*Rating Scale for Level of Comparability

- 1=Weak
- 2=Moderate
- 3=Strong

SAT Reading Descriptor		NAEP Objective(s)	Round 1 Ratings			Round 2 Ratings							
			Overall Coverage		Level of Comparability*			Overall Coverage		Level of Comparability*			
A <i>Determining the Meaning of Words (continued)</i>													
7	Use context clues (such as an embedded definition) to select missing vocabulary at the sentence level		Yes	No	1	2	3		Yes	No	1	2	3
8	Use the context of a sentence or a short section of text to clarify the meaning of unknown words (when definitions may or may not be embedded in the text) or to select the appropriate meaning of familiar and simple words that have multiple meanings)		Yes	No	1	2	3		Yes	No	1	2	3
9	Use sentence structure to negotiate the meaning of the sentence		Yes	No	1	2	3		Yes	No	1	2	3
10	Make sense of complex sentences with logical constructions that include terms such as <i>but, although, or, if, then, and not</i>		Yes	No	1	2	3		Yes	No	1	2	3
11	Use the context of a sentence or larger section of text to determine the meaning of unknown words or to differentiate among multiple possible meanings of words		Yes	No	1	2	3		Yes	No	1	2	3
12	Understand how syntax (the arrangement of words and phrases in a sentence) influences the relationship among words and ideas within a sentence		Yes	No	1	2	3		Yes	No	1	2	3
13	Demonstrate increased comprehension of specialized vocabulary		Yes	No	1	2	3		Yes	No	1	2	3

*Rating Scale for Level of Comparability

- 1=Weak
- 2=Moderate
- 3=Strong

SAT Reading Descriptor		NAEP Objective(s)	Round 1 Ratings			Round 2 Ratings							
			Overall Coverage		Level of Comparability*			Overall Coverage		Level of Comparability*			
A <i>Determining the Meaning of Words (continued)</i>													
14	Understand familiar words in unfamiliar contexts and differentiate among multiple possible meanings for words in unfamiliar contexts		Yes	No	1	2	3		Yes	No	1	2	3
15	Understand sophisticated and specialized vocabulary		Yes	No	1	2	3		Yes	No	1	2	3
16	Determine the meaning of a word when there is little or no supporting context		Yes	No	1	2	3		Yes	No	1	2	3
17	Negotiate complex syntax (the arrangement of words and phrases in a sentence), and integrate ideas within and across sentences		Yes	No	1	2	3		Yes	No	1	2	3
18	Understand how words can sometimes be used in unusual ways that directly refute common usage		Yes	No	1	2	3		Yes	No	1	2	3
19	Access broad and extensive vocabulary within complex syntactical structures and in a variety of contexts		Yes	No	1	2	3		Yes	No	1	2	3
20	Analyze the context of a sentence or larger sections of text to clarify the meaning of unknown words, differentiate among multiple possible meanings of words, detect nuances, and infer connotations		Yes	No	1	2	3		Yes	No	1	2	3

*Rating Scale for Level of Comparability

- 1=Weak
- 2=Moderate
- 3=Strong

SAT Reading Descriptor		NAEP Objective(s)	Round 1 Ratings						Round 2 Ratings					
			Overall Coverage		Level of Comparability*			Overall Coverage		Level of Comparability*				
<i>B Understanding Literary Elements</i>														
1	Identify nuances and attitudes of characters		Yes	No	1	2	3	Yes	No	1	2	3		
2	Determine characterization from dialogue, thoughts and actions, interactions among characters, and narrative perspective		Yes	No	1	2	3	Yes	No	1	2	3		
3	Analyze characters' function in a narrative		Yes	No	1	2	3	Yes	No	1	2	3		
4	Interpret dialogue from a character's or from the narrator's perspective		Yes	No	1	2	3	Yes	No	1	2	3		
5	Analyze the roles and relationships among characters and between character and narrator		Yes	No	1	2	3	Yes	No	1	2	3		

*Rating Scale for Level of Comparability

- 1=Weak
- 2=Moderate
- 3=Strong

SAT Reading Descriptor		NAEP Objective(s)	Round 1 Ratings						Round 2 Ratings					
			Overall Coverage		Level of Comparability*			Overall Coverage		Level of Comparability*				
<i>C Organization and Ideas</i>														
1	Understand the central idea(s) in a simple text or in a short section of a longer text		Yes	No	1	2	3	Yes	No	1	2	3		
2	Determine the main idea of a text and apply it to a different context		Yes	No	1	2	3	Yes	No	1	2	3		
3	Understand the relationship of ideas within and across different texts		Yes	No	1	2	3	Yes	No	1	2	3		
4	Integrate information from short sections of different texts		Yes	No	1	2	3	Yes	No	1	2	3		
5	Integrate ideas within and across sentences and texts		Yes	No	1	2	3	Yes	No	1	2	3		
6	Comprehend generalizations about texts		Yes	No	1	2	3	Yes	No	1	2	3		
7	Analyze and compare concepts across texts		Yes	No	1	2	3	Yes	No	1	2	3		
8	Draw text-based conclusions beyond the main idea		Yes	No	1	2	3	Yes	No	1	2	3		
9	Determine the function of a selected portion of text within a longer text		Yes	No	1	2	3	Yes	No	1	2	3		
10	Analyze main ideas and concepts within and across complex and sometimes opposing texts		Yes	No	1	2	3	Yes	No	1	2	3		
11	Compare and contrast explicit and implicit supporting ideas across texts		Yes	No	1	2	3	Yes	No	1	2	3		

*Rating Scale for Level of Comparability

- 1=Weak
- 2=Moderate
- 3=Strong

SAT Reading Descriptor		NAEP Objective(s)	Round 1 Ratings			Round 2 Ratings							
			Overall Coverage		Level of Comparability*			Overall Coverage		Level of Comparability*			
C		<i>Organization and Ideas (continued)</i>											
12	Recognize components of an author's argument within a text		Yes	No	1	2	3		Yes	No	1	2	3
13	Analyze context, sentence structure, and sentence variation to construct meaning within and across sentences and texts		Yes	No	1	2	3		Yes	No	1	2	3
14	Interpret multiple layers of a text		Yes	No	1	2	3		Yes	No	1	2	3

*Rating Scale for Level of Comparability

- 1=Weak
- 2=Moderate
- 3=Strong

SAT Reading Descriptor		NAEP Objective(s)	Round 1 Ratings						Round 2 Ratings					
			Overall Coverage		Level of Comparability*			Overall Coverage		Level of Comparability*				
<i>D Author's Craft</i>														
1	Use vocabulary clues to determine the tone of a text (the author's attitude toward the subject of the text and toward the audience)		Yes	No	1	2	3		Yes	No	1	2	3	
2	Recognize elements of figurative language (such as metaphor) in a text		Yes	No	1	2	3		Yes	No	1	2	3	
3	Identify an author's purpose for writing		Yes	No	1	2	3		Yes	No	1	2	3	
4	Identify and describe the effects of literary devices used to achieve a specific purpose		Yes	No	1	2	3		Yes	No	1	2	3	
5	Infer the author's opinion concerning the central ideas in a text		Yes	No	1	2	3		Yes	No	1	2	3	
6	Use tone to infer an author's unstated assumptions		Yes	No	1	2	3		Yes	No	1	2	3	
7	Use context clues to identify an author's rhetorical purpose (for example, to persuade the audience) in a short section of text		Yes	No	1	2	3		Yes	No	1	2	3	
8	Infer an author's purpose for writing		Yes	No	1	2	3		Yes	No	1	2	3	
9	Recognize the use of irony and the effects of other sophisticated literary devices, such as symbolism, in a text		Yes	No	1	2	3		Yes	No	1	2	3	
10	Determine the function of words and devices in limited sections of text		Yes	No	1	2	3		Yes	No	1	2	3	
11	Analyze how an author achieves specific effects using rhetorical devices and strategies		Yes	No	1	2	3		Yes	No	1	2	3	

*Rating Scale for Level of Comparability

- 1=Weak
- 2=Moderate
- 3=Strong

SAT Reading Descriptor		NAEP Objective(s)	Round 1 Ratings						Round 2 Ratings					
			Overall Coverage		Level of Comparability*			Overall Coverage		Level of Comparability*				
<i>D Author's Craft (continued)</i>														
12	Analyze an author's explicit and implicit purposes for writing		Yes	No	1	2	3		Yes	No	1	2	3	
13	Analyze the effects of an author's rhetorical and stylistic choices		Yes	No	1	2	3		Yes	No	1	2	3	
14	Distinguish among opinion, fact, conjecture, and hypothesis in a text		Yes	No	1	2	3		Yes	No	1	2	3	
15	Recognize how an author uses evidence to support a particular position		Yes	No	1	2	3		Yes	No	1	2	3	
16	Recognize subtleties and differences in tone, such as the use of humor or irony to achieve a specific effect		Yes	No	1	2	3		Yes	No	1	2	3	
17	Analyze the overall purpose of an author's text		Yes	No	1	2	3		Yes	No	1	2	3	
18	Analyze how sophisticated rhetorical devices support an author's purposes		Yes	No	1	2	3		Yes	No	1	2	3	
19	Analyze how the sophisticated use of literary devices and figurative language (such as extended metaphors, complicated analogies, and symbolic images) achieve specific effects		Yes	No	1	2	3		Yes	No	1	2	3	

*Rating Scale for Level of Comparability

- 1=Weak
- 2=Moderate
- 3=Strong

SAT Reading Descriptor		NAEP Objective(s)	Round 1 Ratings			Round 2 Ratings		
			Overall Coverage	Level of Comparability*		Overall Coverage	Level of Comparability*	
<i>E Reasoning and Inferencing</i>								
1	Perform clear, simple steps of reasoning		Yes	No	1	2	3	Yes No 1 2 3
2	Recognize a general idea, such as a paraphrase, that is supported by separate but related points in different sentences		Yes	No	1	2	3	Yes No 1 2 3
3	Perform multiple steps of reasoning		Yes	No	1	2	3	Yes No 1 2 3
4	Make multiple, layered inferences		Yes	No	1	2	3	Yes No 1 2 3
5	Make layered inferences and apply those inferences to different but related situations		Yes	No	1	2	3	Yes No 1 2 3
6	Make a connection between one part of a text and a later part of that same text to enhance comprehension		Yes	No	1	2	3	Yes No 1 2 3
7	Perform complex reasoning tasks on short sections of text		Yes	No	1	2	3	Yes No 1 2 3
8	Determine an author's unstated assumptions and develop inferences from explicit evidence in different sections of a text		Yes	No	1	2	3	Yes No 1 2 3
9	Draw multiple extended inferences that require several steps of reasoning		Yes	No	1	2	3	Yes No 1 2 3
10	Draw inferences based on implications throughout a text		Yes	No	1	2	3	Yes No 1 2 3
11	Consider the entire text when making inferences, linking information to ideas both before and after a specific section		Yes	No	1	2	3	Yes No 1 2 3

*Rating Scale for Level of Comparability

- 1=Weak
- 2=Moderate
- 3=Strong

SAT Reading Descriptor		NAEP Objective(s)	Round 1 Ratings			Round 2 Ratings						
			Overall Coverage		Level of Comparability*			Overall Coverage		Level of Comparability*		
<i>E Reasoning and Inferencing (continued)</i>												
12	Integrate both general and detailed information across texts		Yes	No	1	2	3	Yes	No	1	2	3
13	Make inferences when there is no explicit reinforcement in the text or when information is missing, and use those inferences to draw further conclusions about the text		Yes	No	1	2	3	Yes	No	1	2	3
14	Apply conclusions drawn from a text to other contexts, understanding similar or analogous situations in the process		Yes	No	1	2	3	Yes	No	1	2	3
15	Identify an author's unstated assumptions and draw further conclusions about the text based on these assumptions		Yes	No	1	2	3	Yes	No	1	2	3
16	Analyze and relate multiple perspectives on similar topics across texts		Yes	No	1	2	3	Yes	No	1	2	3
17	Compare and contrast deeply embedded details or ideas across texts		Yes	No	1	2	3	Yes	No	1	2	3

*Rating Scale for Level of Comparability

- 1=Weak
- 2=Moderate
- 3=Strong

Appendix H

Initial Evaluation/Ready-to-Proceed Form

Panelist #: _____

SAT/NAEP Comparability Study Reading

Initial Evaluation/Ready-to-Proceed Form

The purpose of this evaluation form is to get your feedback about the adequacy of the explanations and preparation you have received in order to make your judgments of content-area comparability.

Please read each statement and place an "X" in the box to represent your response.

Statement	Rating			
	Strongly Agree	Agree	Disagree	Strongly Disagree
I understand the purpose of the study.				
The overview of the assessments was presented clearly.				
The steps that I am to follow to make my content-area ratings were presented clearly.				
I understand what I will be expected to do to complete my content-area ratings.				

I am ready to proceed and to make my first set of content-area comparability judgments.

Yes No

If no, what other information/explanations do you need before making your first set of judgments?

Date _____ Signature _____

Appendix I
Overall Rating Form

Panelist #: _____

**NAEP/SAT Comparability Study
Mathematics**

Panelist Rating Form—Overall Ratings

Rating 1—Overall Level of Content Comparability

Please indicate below the degree to which you agree with the following statement:

Based on the descriptors and objectives, the content of the SAT and NAEP are comparable.

Strongly Agree
1

Agree
2

Disagree
3

Strongly Disagree
4

Rating 2—Overall Breadth of Content Coverage

Please indicate below the degree to which you agree with the following statement:

Based on the descriptors and objectives, the overall breadth of the SAT and NAEP are comparable.

Strongly Agree
1

Agree
2

Disagree
3

Strongly Disagree
4

Rating 3—Advisability of Future Study

Please indicate below whether you agree with the following statement:

There is sufficient overall overlap between the SAT and NAEP to justify conducting a more extensive alignment study at the item level.

Yes

No

Appendix J

Final Evaluation Form

Panelist #: _____

SATNAEP Comparability Study Reading

Final Evaluation Form

The purpose of this evaluation form is to get your feedback about the overall study.

Please read each statement and place an "X" in the box to represent your response.

Statement	Rating			
	Strongly Agree	Agree	Disagree	Strongly Disagree
The SAT and NAEP skill statements were sufficiently detailed to judge comparability.				
<i>Content-Area Comparability Ratings</i>				
The content-area rating form was easy to complete.				
The summary of our content-area ratings was presented clearly.				
The discussion of the summary of content-area ratings was informative.				
The process of completing the content-area ratings was easy to follow.				
<i>Overall Ratings</i>				
The overall rating form was easy to complete.				
The summary of our overall ratings was presented clearly.				
The discussion of the summary of overall ratings was informative.				
The process of completing the overall ratings was easy to follow.				

Panelist #: _____

What we should consider doing differently the next time we conduct this type of study?

Thank you!