A COGNITIVE LAB REPORT FOR THE
AMERICAN DIPLOMA PROJECT
ALGEBRA I END-OF-COURSE EXAM

Test, Measurement & Research Services

October 14, 2010
Abstract

In an effort to better understand high non-response rates and overall low performance on the American Diploma Project (ADP) Algebra I End-of-Course Exam, a cognitive lab was conducted that allowed for investigation of students’ familiarity with exam items as well as the strategies that students are engaging in when attempting to solve the problems. Data collected through the one-on-one student interviews of 18 students provided an in-depth look at how students are interacting with exam items. The structure of the analyses was exploratory in nature and provided insight as to where in the problem solving process students are struggling, for example, in the selection of appropriate strategies versus the execution of the strategies. Results suggest that students may not be acquiring content knowledge at the depth that is required of the items appearing on the ADP Algebra I End-of-Course Exam. In addition, students may be unfamiliar with demonstrating knowledge through constructed-response items, particularly with extended-response items. The results of the study should be interpreted with a consideration of the study limitations, which include the relatively small samples of students and items that were utilized in the data collection.
American Diploma Project
Algebra I End-of-Course Exam Cognitive Lab

To investigate the basis of the high non-response rate for constructed-response items and the overall low performance for students taking the ADP Algebra I End-of-Course Exam, a cognitive lab study was conducted. Through a series of one-on-one student interviews, Pearson content specialists asked students to “think out loud” when looking at and solving ADP Algebra I Exam items. The structure of the analyses was exploratory in nature and aimed at gaining insight as to where in the problem solving process students are struggling, for example, in the selection of appropriate strategies versus the execution of the strategies. The goals of the study were to:

- Better understand students’ familiarity and comfort level with the range of content on the ADP Algebra I End-of-Course Exam
- Investigate students’ familiarity with exam items and to better understand the problem-solving strategies that students apply

**Framework**

Assessment developers and researchers have been expanding their investigations of the validity and accessibility of tests and items. A particular area of focus has utilized methodologies for investigating student experiences during testing situations while interacting with particular items, and their approaches to problem solving. Many of these investigations have been based on the cognitive lab methodology, originally developed in 1945 by Karl Duncker (Campbell, 1999; Desimone & Le Floch, 2004; Paulsen & Levine, 1999). In order to investigate particular features of item and test validity, and to fully understand the factors impacting student problem solving the think
aloud process has recently been augmented by purposeful probing with targeted questions. Verbal protocols and protocol analysis techniques (Ericsson & Simon, 1993) are the basis of extracting useful information from data collected using the cognitive interviewing methodology. Evaluating cognitive processes, such as students’ comprehension and problem solving techniques, requires a qualitative, data-rich methodology, such as that provided by cognitive interviews, or “think-alouds” (Ericsson & Simon, 1993; Leighton, 2004). Verbal reports offer an important tool in examining how algebra students approach items, especially extended-response items, because they provide different, more direct evidence than methods such as observation and surveys.

Methods

Sample

Because New Jersey is the only state to administer the ADP Algebra I End-of-Course Exam to all of its Algebra I students, students from New Jersey were selected to participate in the cognitive labs. Pearson requested that the state select three schools for the study. The criterion for selection was at the states’ discretion and was contingent on schools being able to participate in the study during the designated timeframe of June 7-10. The State Department of Education allowed Pearson to offer participating schools $50 per student for participation in the study so all three schools that participated in the study received $300.

Each of the three schools that participated was asked to identify six students: two high performing students, two middle performing students, and two lower performing students. However, one school identified two lower performing, one middle performing, and three high performing students resulting in an unbalanced design. The level of
performance was to be based on an external variable such as current GPA, anticipated grade in current mathematics course, standardized test scores, or other criteria, at the schools’ discretion to select students. The intent was to gather information from students across a range of achievement levels both within and across schools, although there was no attempt made to standardize the definitions for low, middle, and high performers. The achievement level of the students was not shared with the interviewers in advance but was used during the research analysis portion of the study.

Participating students took the ADP Algebra I spring 2010 exam approximately two weeks before they participated in the study. The timing of the cognitive labs was somewhat restricted by external factors. The labs had to be conducted during the school year in order to ensure access to students at the time when they have been exposed to the content in Algebra I classes. Additionally, the labs had to occur after students had participated in the ADP Algebra I assessment so that the lab experience would not influence their operational score since the cognitive lab included operational items.

Pearson provided an informational overview and parental permission form to obtain consent for minors to participate in the study. The forms were also provided in Spanish upon request from one of the participating schools. All students turned in completed permission slips on the day of the labs.

**Data Collection**

The cognitive labs were conducted as one-on-one interview sessions. Each student session was scheduled for 90 minutes and was led by a Pearson mathematics content specialist who served as the interviewer. A total of 18 Algebra I students participated. All sessions were conducted in person at the students’ schools during school
hours the week of June 7, 2010. Interviews were audio recorded and all student work was collected.

Prior to conducting the cognitive labs Pearson developed a protocol and script that the interviewers used to guide the sessions and to elicit student responses as they were shown ADP Algebra I Exam items. The items selected for the study included 12 released items and 6 spring 2010 operational items with operational items being selected on the basis of having high omit rates. (Sample items can be found at http://education.pearsonassessments.com/pai/ea/state/end+of+course+assessment.htm.) The items included both calculator and non-calculator items and students were instructed to bring their calculator to the study. Students were shown a mixture of item types representative of those on the exam—multiple-choice, short-response, and extended-response—that measured the four content standards: Operations on Numbers and Expressions (O), Linear Relationships (L), Non-linear Relationships (N), and Data, Statistics, and Probability (D). (See Appendix A for a list of benchmarks.)

For the released items, two multiple-choice items and one constructed-response item were selected from each of the four content standards. Students were shown each item (without actually working the problem) and asked if 1) they were familiar with the content of the item and 2) they would likely get the question correct, partially correct (for selected-response items this means they could eliminate one or two answer choices), or incorrect.

For the “think-aloud” portion, students worked through two multiple-choice (MC) items worth one-point each, three short-answer (SA) items worth two points each, and one-extended response (ER) item worth four points. All of these items appeared in the
spring 2010 operational exam. It should be noted that because of security considerations, neither the items nor detailed transcripts of the students’ think-aloud protocols can be shared in this report. Students were asked to solve the items while talking out loud to the interviewer about what they were thinking. Interviewers assured students that their performance would not be scored but that it was the student’s problem solving process that was being studied. The conversation was primarily led by the student’s thinking aloud as he or she worked through the problems. The script provided interviewers with types of questions to ask students if they were not verbalizing their strategies.

After viewing released and operational items, interviewers asked students about two additional issues, affect and motivation. Students were asked about their overall feelings regarding participation in the cognitive lab (dreaded, didn’t care, interested) and their relative nervousness, anxiety, and motivation on the cognitive lab compared to the operational administration of the exam.

Analyses

Coding

Students’ verbal responses were recorded and transcribed. Then using transcriptions, audio files, and hard copies of the students’ work, content specialists coded each student’s response. Coding was completed by two content specialists. After training, 20% double coding was done to address quality assurance and reliability. The agreement standard was 75% for all coding with the exception of two coding categories: the appropriateness of a selected strategy and the execution within that strategy. Due to the nested nature of these two coding categories, a higher standard of 85% agreement was required. Whenever the agreement standard was not met, content specialists met to
further discuss the coding process and resolve discrepancies. The coding of each interview was recorded in an Excel workbook that contained three sheets, one for each of the following sections.

**Familiarity and estimated performance on released items.** The first sheet allowed for coding the familiarity (familiar, unfamiliar, no response) and performance estimate (correct, partially correct, incorrect) for each item. Data were recorded by checking the appropriate box.

**Strategies implemented on operational items.** Content specialists created coding schemes by identifying possible problem-solving strategies that students may attempt while working to solve each item. A unique list of strategies was created for each item and contained in a drop-down menu in the second sheet of the Excel workbook. Content specialists could then identify the strategies that a student applied to each item and then indicate whether the selected strategy was appropriate and whether that strategy was executed correctly. An indication was also made as to whether the student correctly solved each item.

**Affect and motivation questions.** The third sheet provided a place to record the affect and motivation responses regarding the cognitive lab experience. Responses were recorded by checking the appropriate box.

**Frequencies**

Once coding was entered in the Excel workbooks, data were read into SAS for analysis. Due to the exploratory nature of this study, numerous frequencies were calculated exploring the various factors of the design including achievement level, item type, strategy appropriateness, strategy execution, and item performance.
Results

Familiarity and Estimated Performance on Released Items

Table 1 displays students’ familiarity with the items and their estimated performance. The familiarity column reports the percentage of students who indicated that they were familiar with the content that the item measures. For estimated performance, the percentages of students who estimated that they would get the item correct, partially correct, and incorrect are displayed. Partially correct was defined as getting some of the points on the constructed-response items and being able to eliminate one or two distractors on multiple-choice items. Across item type and standard, at least 72% of students indicated familiarity with all of the items. For estimated performance, the percentage of students who felt confident that they could get the item correct ranged from 17% to 100% across the 12 items. Interestingly, the percentages of students who estimated that they would receive full credit on the constructed-response items was over 60%, but in reality students consistently perform poorly on the constructed-response items.
Table 1. Familiarity and Estimated Performance on Released Items (n=18)

<table>
<thead>
<tr>
<th>Item</th>
<th>Benchmark</th>
<th>Calc/No Calc</th>
<th>Item Type</th>
<th>Familiarity</th>
<th>Correct</th>
<th>Partial</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O,1,c</td>
<td>No Calc</td>
<td>MC</td>
<td>72%</td>
<td>17%</td>
<td>56%</td>
<td>28%</td>
</tr>
<tr>
<td>2</td>
<td>L,1,c</td>
<td>No Calc</td>
<td>MC</td>
<td>100%</td>
<td>44%</td>
<td>44%</td>
<td>11%</td>
</tr>
<tr>
<td>3</td>
<td>N,1,a</td>
<td>No Calc</td>
<td>MC</td>
<td>83%</td>
<td>50%</td>
<td>28%</td>
<td>22%</td>
</tr>
<tr>
<td>4</td>
<td>D,2,b</td>
<td>No Calc</td>
<td>MC</td>
<td>100%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>5</td>
<td>O,2,d</td>
<td>Calc</td>
<td>MC</td>
<td>72%</td>
<td>33%</td>
<td>28%</td>
<td>39%</td>
</tr>
<tr>
<td>6</td>
<td>L,2,e</td>
<td>Calc</td>
<td>MC</td>
<td>100%</td>
<td>83%</td>
<td>11%</td>
<td>6%</td>
</tr>
<tr>
<td>7</td>
<td>N,2,a</td>
<td>Calc</td>
<td>MC</td>
<td>72%</td>
<td>39%</td>
<td>39%</td>
<td>22%</td>
</tr>
<tr>
<td>8</td>
<td>D,1,b</td>
<td>Calc</td>
<td>MC</td>
<td>100%</td>
<td>89%</td>
<td>6%</td>
<td>6%</td>
</tr>
<tr>
<td>9</td>
<td>L,1,d</td>
<td>No Calc</td>
<td>SA</td>
<td>89%</td>
<td>61%</td>
<td>31%</td>
<td>8%</td>
</tr>
<tr>
<td>10</td>
<td>D,1,a</td>
<td>No Calc</td>
<td>SA</td>
<td>89%</td>
<td>72%</td>
<td>22%</td>
<td>6%</td>
</tr>
<tr>
<td>11</td>
<td>N,2,b</td>
<td>Calc</td>
<td>SA</td>
<td>100%</td>
<td>89%</td>
<td>6%</td>
<td>6%</td>
</tr>
<tr>
<td>12</td>
<td>O</td>
<td>No Calc</td>
<td>ER</td>
<td>93%</td>
<td>65%</td>
<td>15%</td>
<td>20%</td>
</tr>
</tbody>
</table>

Strategies Implemented on Operational Items

The six operational items (two MC, three SA, and one ER) that comprised the think-aloud portion of the study proved to be very difficult for students regardless of achievement level. The average number correct for each of the lower, middle, and high achievement levels were 0.5, 0.6, and 1.1 respectively. Nine students were unable to answer any of the items correctly and the top performing students only answered two correctly. Although there was little difference in overall item performance across the three achievement levels, there were differences in the number of strategies that students engaged in, whether students selected appropriate strategies, and whether the strategies were executed correctly.

Non-attempts. There were some instances when students did not even make an attempt to solve an item. Although interviewers prompted students in these situations, some students were at a complete loss as to how to even begin some items. Table 2 indicates the number of items skipped completely for each item type across the six students that participated at each achievement level. Keep in mind that the six item set
was comprised of two multiple-choice items, three short-answer items, and one extended-
response item so two students skipping the extended-response item eliminates strategy
data for one-third of the lower achievement level. In addition, having eight short-answer
items unattempted eliminates 15% of the data for the short-answer item type. Across
achievement levels, students in the lower achievement level skipped considerably more
items (n=7) than students in the other achievement levels.

Table 2. Non-Attempted Items

<table>
<thead>
<tr>
<th></th>
<th>MC</th>
<th>SA</th>
<th>ER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower</td>
<td>0</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Medium</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>High</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

**Number of strategies.** The average number of strategies that students engaged in
throughout the entire set of six items increased with achievement level from 13 to 17 to
19 with the distribution across item types varying. Figure 1 shows an interesting pattern
by achievement level and item type. It is intuitive that the number of strategies engaged
in for an item would increase across the item types of multiple-choice, short-answer, and
extended-response. While this pattern held for multiple-choice relative to short-answer
items, students applied the fewest strategies to the extended-response item. In addition,
the number of strategies applied tended to increase between the lower and medium
achievement levels and a sharp increase was seen in the short-answer strategies between
medium and high achievement levels.
Figure 1. Average Number of Strategies Engaged In

![Average Number of Strategies Engaged In](image)

**Appropriateness of strategies.** Figure 2 displays the percentage of strategies that were appropriate by achievement level and item type. Of the strategies selected, the percentage of appropriate strategies tended to increase with achievement level, particularly so for short-answer items. For the multiple-choice and extended-response items, roughly 80% of the strategies that students engaged in were appropriate problem-solving strategies.
Execution of strategies. Regardless of whether the strategy that a student engaged in was appropriate, the percentages of correct execution by achievement level and item type were calculated and are shown in Figure 3. No general trends appear across the achievement levels. However, strategies for extended-response times were rarely executed properly while strategies for multiple-choice items were correctly executed most of the time.
Affect and Motivation Questions

Of the 18 participants, 13 indicated that they were interested in participating in the cognitive lab. In general, students indicated feeling less nervous and anxious on the cognitive lab compared to the operational administration. However, students were split fairly evenly regarding whether they were more or less motivated for the cognitive lab compared to the operational administration.

Discussion

The first portion of this study examined students’ familiarity with ADP Algebra I Exam items but the majority of the focus was on how achievement level and item type affect students’ problem solving strategies. Several interesting discussion points emerged: 1) familiarity and high estimated performance on released items, 2) difficulty of all six operational items for all achievement levels, and 3) strategy trends within item types.
Familiarity and High Estimated Performance

When shown the twelve released items, regardless of achievement level and item type, students tended to indicate high familiarity with the content. In addition, students estimated high performance on many of the items including all constructed-response items although this is contrary to how students actually perform. While these results could be attributed to the tendency for self-reported data to be overestimated, another explanation is that students are familiar with content but not to the depth at which the ADP Algebra I Exam items require. In addition, students may have difficulty demonstrating their knowledge through short-answer and extended-response items. Although students indicate high estimated performance on the constructed-response items, perhaps they don’t fully understand the steps that are needed in order to answer them correctly.

Difficulty of items

In the think-aloud portion of the study when students actually worked through the six operational items, few items were answered correctly. The average number of correct answers was 1 out of 6 for all three achievement levels. Further indication of the difficulty of the constructed-response items was the frequency with which students were unable even to attempt an item. While this could possibly be a function of item type, students were also unable to correctly answer multiple-choice items suggesting that students don’t possess the depth of content knowledge required on the ADP Algebra I Exam.
**Strategy Trends by Item Type**

When interpreting the data regarding strategies, it is important to keep in mind that the number of strategies required to correctly answer an item varies by item and student. An item that takes Student A one step to answer may take another Student B three steps, depending on how they go about solving the problem, and they may both arrive at the correct answer. Additionally, constructed-response items are more demanding than multiple-choice items which may translate into constructed-response items requiring more complex strategies in order to arrive at a correct answer.

**MC Items.** Although students were often times unable to correctly answer the multiple-choice items, they were able to select appropriate strategies roughly 80% of the time and also found success in execution of strategies. However, it appears that while students could make progress on multiple-choice items, they were unable to meet the demands of the item in order to arrive at the correct answer. This result could suggest that although students may be familiar with the content and with responding to multiple-choice items, the ADP Algebra I Exam items may be requiring a deeper level of understanding.

**SA Items.** Across achievement levels students tended to engage in more strategies on short-answer items as compared to multiple-choice items with the high achievement group engaging in roughly twice as many strategies. However, for students in the lower and medium achievement levels, often times those strategies were not appropriate nor executed correctly. The lower and medium achievement levels selected appropriate strategies only half of the time and with correct execution of strategies less than 50%. But strategy selection by the high achievement level students was appropriate around three-
fourths of the time and 87% of strategies were executed correctly. This contrast in strategy selection and proper execution between achievement levels could be an indication that short-answer items demand more than the average student is currently capable of doing.

**ER Items.** Data suggest that students, regardless of achievement level, are at a loss for how to respond in an extended-response format. Although it is intuitive that these items would require more strategies, students engaged in the fewest strategies with this item type suggesting that students may be unable to understand what the item is asking of them. Figure 2 indicates high percentages of selected strategies being appropriate for extended-response items and Figure 3 shows relatively low correct execution percentages. Therefore, it appears that although students were unable to make much progress on the extended-response item, the strategies that students engaged in were often times appropriate although students often failed to correctly execute the strategies. Potential explanations for these results include that students may be unfamiliar with either or both content and item type.

**Limitations and Conclusions**

This cognitive lab study was an exploratory study that allowed for an in-depth investigation of students’ familiarity with ADP Algebra I Exam items as well as the strategies that students are engaging in when attempting to solve the problems. However, several limitations should be taken into account when interpreting results. First, the released items have been used extensively in professional development so students could be accurate in their responses regarding familiarity and expected performance if they have been used in the classroom. Second, there is a trade off in this type of study of being
able to investigate fewer items and students in exchange for more in-depth data. The items included in this study were few, measured various standards, had different calculator requirements, and were of varying item types. Furthermore, only between five and seven students made up each of the three achievement levels and these students were from three schools in one state causing problems for generalizations. All of these issues should be considered when interpreting the data.

Despite these limitations, the cognitive lab study provided rich data and shed light on potential reasons why students find difficulty in the ADP Algebra I Exam items. First, students may not be acquiring content knowledge at the depth that is required from the items. Some students were able to engage in appropriate strategies and correctly execute them but still were not able to arrive at the correct answer. Second, students may be unfamiliar with demonstrating knowledge through constructed-response items. In general, students were unable to identify appropriate strategies for extended-response items and when they could, found difficulty with correct execution. This also held true for the lower and medium achievement levels for short-answer items but students in the high achievement level were more successful in selection appropriate strategies and proper execution. The inability for many students to determine an appropriate strategy for constructed-response items may be leading them to skip the items altogether, resulting in the high omit rates that have been observed.
References


## ADP Algebra I End-of-Course Exam Content Standards

<table>
<thead>
<tr>
<th>Content Standards</th>
<th>Content Benchmarks</th>
</tr>
</thead>
</table>
| **O1. Number Sense and Operations** | a. Use properties of number systems within the set of real numbers to verify or refute conjectures or justify reasoning and to classify, order, and compare real numbers.  
b. Use rates, ratios and proportions to solve problems, including measurement problems.  
c. Apply the laws of exponents to numerical expressions with integral exponents to rewrite them in different but equivalent forms or to solve problems.  
d. Use the properties of radicals to rewrite numerical expressions containing square roots in different but equivalent forms or to solve problems. |
| **O2. Algebraic Expressions** | a. Apply the laws of exponents to algebraic expressions with integral exponents to rewrite them in different but equivalent forms or to solve problems.  
b. Add, subtract and multiply polynomial expressions with or without a context.  
c. Factor simple polynomial expressions with or without a context.  
d. Use the properties of radicals to convert algebraic expressions containing square roots into different but equivalent forms or to solve problems. |
| **L1. Linear Functions** | a. Recognize, describe and represent linear relationships using words, tables, numerical patterns, graphs and equations. Translate among these representations.  
b. Describe, analyze and use key characteristics of linear functions and their graphs.  
c. Graph the absolute value of a linear function and determine and analyze its key characteristics.  
d. Recognize, express and solve problems that can be modeled using linear functions. Interpret their solutions in terms of the context of the problem. |
| **L2. Linear Equations and Inequalities** | a. Solve single-variable linear equations and inequalities with rational coefficients.  
b. Solve equations involving the absolute value of a linear expression.  
c. Graph and analyze the graph of the solution set of a two-variable linear inequality.  
d. Solve systems of linear equations in two variables using algebraic and graphic procedures.  
e. Recognize, express and solve problems that can be modeled using single-variable linear equations; one- or two-variable inequalities; or two-variable systems of linear equations. Interpret their solutions in terms of the context of the problem. |
| **N1. Non-linear Functions** | a. Recognize, describe, represent and analyze a quadratic function using words, tables, graphs or equations.  
b. Analyze a table, numerical pattern, graph, equation or context to determine whether a linear, quadratic or exponential relationship could be represented. Or, given the type of relationship, determine elements of the table, numerical pattern or graph.  
c. Recognize and solve problems that can be modeled using a quadratic function. Interpret the solution in terms of the context of the original problem. |
| **N2. Non-linear Equations** | a. Solve equations involving several variables for one variable in terms of the others.  
b. Solve single-variable quadratic equations. |
| **D1: Data and Statistical Analysis** | a. Interpret and compare linear models for data that exhibit a linear trend including contextual problems.  
b. Use measures of center and spread to compare and analyze data sets.  
c. Evaluate the reliability of reports based on data published in the media. |
| **D2: Probability** | a. Use counting principles to determine the number of ways an event can occur. Interpret and justify solutions.  
b. Apply probability concepts to determine the likelihood an event will occur in practical situations. |