An Investigation of the Test-Taking Skills and Associated Thinking Processes of Ninth Grade English Language Arts Students

Mary B. O'Shea EdD

University of San Diego

Follow this and additional works at: https://digital.sandiego.edu/dissertations

Part of the Leadership Studies Commons

Digital USD Citation
https://digital.sandiego.edu/dissertations/902

This Dissertation: Open Access is brought to you for free and open access by the Theses and Dissertations at Digital USD. It has been accepted for inclusion in Dissertations by an authorized administrator of Digital USD. For more information, please contact digital@sandiego.edu.
AN INVESTIGATION OF THE TEST-TAKING SKILLS AND ASSOCIATED THINKING PROCESSES OF NINTH GRADE ENGLISH LANGUAGE ARTS STUDENTS

by

Mary B. O’Shea

A dissertation submitted to the faculty of San Diego State University and the University of San Diego in Partial Fulfillment of the Requirements for the Degree Doctor of Education

Dissertation Committee:

Barbara Moss, Ph.D., San Diego State University
Marcie Bober-Michel, Ph.D., San Diego State University
Heather Lattimer, Ed.D., University of San Diego

Summer 2010
DEDICATION

This dissertation is dedicated to my late father, Vincent John O'Shea, who told me once that, “an education should never give you any reason to feel superior to others but it should give you every reason not to feel inferior.” After spending nearly every weekend for the past two years toiling through endless articles, writing multiple drafts and analyzing hundreds of pages of data, I often questioned those words as I felt a constant burden inside. Clearly, I didn’t realize the emotional “heaviness” I would face as I meandered through the overwhelming challenges of writing a dissertation while working a full-time teaching job and caring for a precious baby. Time after time, as I saw my beautiful daughter go into the hands of a sitter, I felt unenthused and at the mercy of my computer. There were moments when I felt like giving up but then I thought about what my father had said. I knew I wouldn’t feel overwhelmed forever as long as I finished and honored my commitment to the educational world. I still don’t quite understand what made me undertake such an arduous task except that I had a compelling interest to capture students’ thinking for others to see. Like a soldier in the trenches, I am a teacher that sees many cracks and crevices in the educational landscape. For years, I thought about sharing my insights with others who oversee the battlefield. But then again, I didn’t realize until I embarked on this journey that my students’ insights are the most important. They are the real heroes; those who endure day after day without any ability to truly fight. It’s official now! I am a researcher. Thanks again Dad! I know you are looking down on me.
ABSTRACT OF THE DISSERTATION

An Investigation of the Test-Taking Skills and Associated Thinking Processes of Ninth Grade English Language Arts Students

by

Mary B. O’Shea

Doctor of Education

San Diego State University and the University of San Diego, 2010

Although much is known about how students perform on standardized tests, little research exists concerning how students think and process while taking such tests. This mixed methods action research study was designed to investigate if a constructivist approach to test preparation could yield improved results for 37 English language arts freshmen preparing for the California High School Exit Exam (CAHSEE). The researcher explored student performances on Princeton Review practice CAHSEE tests (versions A, B, and C) over a 3-month period and she performed several statistical tests including repeated measures ANOVA by individual classes (advanced, basic, below basic) to determine differences between and among groups. In addition, she examined students’ metacognitive processes as they participated in bi-monthly audio-taped classroom discussions of their thinking on specific multiple-choice items. After each of these discussions, all students independently wrote to a common prompt in their journals to articulate their thinking. In addition, six focus students recorded their thinking aloud while they tested on specific items and participated in end of study interviews about multiple aspects of the intervention. Emerging themes included how students read and interpreted test questions, eliminated answers, revisited the text, and felt about the intervention. While repeated measures ANOVA indicated a significant testing effect ($p = .05$) for advanced students early in the study, it was the below basic test-takers who actually made the greatest gains overall. And although students made improvement interpreting test questions and eliminating answers, all groups struggled with knowing how to strategically revisit the text. Approximately one-third of the students from all classes attributed laziness or tiredness as a factor influencing their performance. The more advanced focus students tended to respond to the test items in a cyclical pattern, practicing strategies repeatedly whereas the less advanced test-takers participated in more of a lock step, linear approach. Most focus students declared that the most useful aspect of the intervention was thinking aloud as they tested or listening to others during classroom discussions. Recommendations for future studies include more in-depth, multidimensional studies to capture students’ thought processes in order to shape classroom instruction that assists students in becoming better test-takers.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td></td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td></td>
<td>xii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td></td>
<td>xiii</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Purpose of the Study</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>Research Questions</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Theoretical Framework</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Significance of the Study</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Limitations of the Study</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>Definition of Key Terms</td>
<td></td>
<td>14</td>
</tr>
<tr>
<td>2 REVIEW OF LITERATURE</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>The Effects of the <em>No Child Left Behind Act</em> (NCLB)</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>The Evolution of Standardized Testing in America</td>
<td></td>
<td>19</td>
</tr>
<tr>
<td>The Effects of Standardization</td>
<td></td>
<td>22</td>
</tr>
<tr>
<td>Effective Test Preparation</td>
<td></td>
<td>27</td>
</tr>
<tr>
<td>Metacognition</td>
<td></td>
<td>35</td>
</tr>
<tr>
<td>Metacognition and the Classroom</td>
<td></td>
<td>36</td>
</tr>
<tr>
<td>Expert vs. Novice Metacognition</td>
<td></td>
<td>38</td>
</tr>
</tbody>
</table>
3 METHODOLOGY ................................................................. 45

Rationale for Method ...................................................... 45
Research Questions ....................................................... 47
Background/Pilot Work .................................................... 48
The Site and the Participants ............................................ 50
Data Collection .............................................................. 53

Instruments ................................................................. 53

Informed Consent Forms .................................................. 53
Princeton Review Tests ................................................... 53
Classroom Discussions .................................................. 54
Journal Reflections ......................................................... 54
Focus Students Participating in Think Alouds ....................... 54
End of Study Interviews .................................................. 55

Procedures ................................................................. 55

Informed Consent Forms .................................................. 55
Princeton Review Tests ................................................... 56
Classroom Discussions .................................................. 57
Journal Reflections ......................................................... 60
Focus Students Participating in Think Alouds ....................... 62
End of Study Interviews .................................................. 63
4 RESULTS

Introduction to the Quantitative Test Results

Overall Mean Test Results for Advanced, Basic, and Below Basic Classes

Individual Class Mean Test Results on Each Test Over the Course of the Study

Advanced Mean and Standard Deviation Scores Over Time

Basic Mean and Standard Deviation Scores Over Time

Below Basic Mean and Standard Deviation Scores Over Time

Repeated Measures ANOVA: Pairwise Comparisons of the Advanced Group

Repeated Measures ANOVA: Pairwise Comparisons of the Basic Group

Repeated Measures ANOVA: Pairwise Comparisons of the Below Basic Group

Summary of the Quantitative Data

Introduction to the Qualitative Test Results

Qualitative Results

Reading the Question and Answer General Results

March Findings

April Findings

May Findings
5 INTRODUCTION TO THE CONCLUSION ................................. 167

The Problem ......................................................... 167
Purpose of the Study ............................................... 168
Rationale for Method ............................................... 169
Research Questions .................................................. 169
Key Findings .......................................................... 170
The Importance of Examining the Way Students Think and Process .... 171
Interpreting the Test Question ....................................... 172
Recognizing Key Words and "Correct" Answers .................... 173
Revisiting the Text .................................................. 174
The Differences Between High and Low Test-Takers: Cyclical vs. Linear .................................................. 174
The Feelings that Students Have Regarding Test-Taking ............ 176
Putting It All Together: What Made the Lowest Performing Group Improve the Most? ............................................. 176
The Focus Students' Feelings and Preferences ...................... 177
Discussion ............................................................ 179
Not All Low Achievers Lack Motivation ............................ 180
Lack of Motivation May Be Beyond the Teacher's Control .......... 180
Teaching Test Wiseness Is Not Teaching to the Test ............... 181
The Way Test Scores Are Reported May Make a Difference .......... 181
Limitations ........................................................................ 182
Recommendations for Instructional Practice ....................... 183
Recommendations for Future Research ................................. 186
Final Remarks ..................................................................... 187
REFERENCES ..................................................................... 190
APPENDICES

A PROTECTION/ANONYMITY OF SUBJECTS: PARENT
PERMISSIONS LETTER .......................................................... 201

B PROTECTION/ANONYMITY OF SUBJECTS: STUDENT
PERMISSIONS LETTER .......................................................... 206

C PRINCETON REVIEW PRACTICE CAHSEE TEST FORMATS
FOR THIS STUDY ................................................................ 213

D INDIVIDUAL JOURNAL RESPONSE SHEET (HYPOTHETICAL) ...... 218

E INTERVIEW QUESTIONS TO BE ASKED TO FOCUS STUDENTS
AT THE COMPLETION OF THE INTERVENTION .................... 220

F SAMPLE MATRIX FOR ITEM ANALYSIS (HYPOTHETICAL) ........ 223

G GROUP DISCUSSION TRANSCRIPT SAMPLE QUESTIONS
(HYPOTHETICAL) ................................................................. 225
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Constructive Controversy, Debate, and Concurrence Seeking</td>
<td>42</td>
</tr>
<tr>
<td>2</td>
<td>The Spiral Series of Thought</td>
<td>47</td>
</tr>
<tr>
<td>3</td>
<td>Ethnic/Racial Academic Performance Index Subgroup Information</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>in the Small School Where This Study Took Place</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Demographics of Focus Students</td>
<td>63</td>
</tr>
<tr>
<td>5</td>
<td>Mean, Standard Deviation, and Range for All Three PR Practice Tests</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>(Combined)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Mean and Standard Deviation Scores of Individual Classes Over Time</td>
<td>72</td>
</tr>
<tr>
<td>7</td>
<td>Mean and Standard Deviation Scores of the Advanced Class Over Time</td>
<td>73</td>
</tr>
<tr>
<td>8</td>
<td>Focus Student Results on Overall Practice CAHSEE Tests</td>
<td>74</td>
</tr>
<tr>
<td>9</td>
<td>Mean and Standard Deviation Scores of the Basic Class Over Time</td>
<td>75</td>
</tr>
<tr>
<td>10</td>
<td>Mean and Standard Deviation Scores of the Below Basic Class Over Time</td>
<td>76</td>
</tr>
<tr>
<td>11</td>
<td>Pairwise Comparisons of the Advanced Class</td>
<td>77</td>
</tr>
<tr>
<td>12</td>
<td>Pairwise Comparisons of the Basic Class</td>
<td>78</td>
</tr>
<tr>
<td>13</td>
<td>Pairwise Comparisons of the Below Basic Class</td>
<td>80</td>
</tr>
<tr>
<td>14</td>
<td>Continuum of Themed Behaviors Around Test Questions</td>
<td>83</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1. The cycle model: Testing, discussing, reflecting, discussing, reflecting, testing .................................................. 58
CHAPTER 1

INTRODUCTION

This is a mixed methods, action research study of an intervention designed to assist 37 freshmen English students as they prepare to take the California High School Exit Exam (CAHSEE), a graduation requirement in the state of California. Their first official attempt to pass this exam will be in the month of March during their sophomore year, 18 months after they begin high school. This study will include an intervention in which students take three, miniature practice Princeton Review English Language Arts CAHSEE exams (Princeton Review, 2005), consisting of 25 questions each, during a semester segment of the freshman school year. In between practice tests and through whole class discussions, students will discuss challenging test questions in-depth, and then individually reflect on specific test items through journal entries. During the in-class discussions and journal reflections, the study will investigate the metacognitive processes (metacomprehension, problem solving, critical thinking) that students elicit when taking English language arts multiple-choice practice questions. The intervention will also include an additional focus on six students consisting of an in-depth analysis of their processes during testing in which they participate in three recorded think alouds. Half of these focus students will be historically strong test-takers and half weak and there will be a mix of male vs. female. At the conclusion of the study, these focus students will participate in individual interviews in which they share what aspects of the study each felt was the least and most effective and why.
The proposed study emerges from a need for embedded test preparation in high school English language arts classrooms (Crocker, 2005; Jones, Jones, & Hargrove, 2003) as well as a need for robust student-centered test preparation in which students have opportunities to share and discuss their thinking about test items and to reflect on their thinking through written journal responses.

The practice test-taking phase will explore whether differences in performances can be attributed to class level (advanced vs. basic, basic vs. below basic, advanced vs. below basis). Interestingly, research in the past few decades has revealed that males are performing considerably lower than females not only on reading tests (Mullis, Dossey, Foertsch, Jones, & Gentile, 1991), but on writing assessments (Rubin & Greene, 1992). Recent CAHSEE data shows that 10th grade females, on their initial attempts to pass the CAHSEE, considerably outperformed males on the English language arts section (82.9% female pass rate vs. 74.9% male pass rate) and slightly outperformed males on the math sections of the CAHSEE (70.6% female pass rate vs. 70.0% male pass rate; California Department of Education [CDE], 2008). Not only have recent test scores suggested a surge in female scores and a decline in male scores, it has also been established that students with high GPAs consistently outperform those with low GPAs on tests (Hong, Sas, & Sas, 2006; Kim & Goetz, 1993; McClain, 1983). Thus, those in advanced English may inevitably outperform those in basic and below basic English when comparing average mean scores.

While test data typically is reported from year to year comparing differing cohorts, there is a need to track individual students’ progress over time (Darling-Hammond, 2006). Currently, little research exists to see if classes (advanced vs. basic, basic vs. below basic, advanced vs. below basic) outperform others on standardized tests. This study attempts to
track such progress through a series of practice tests. In addition, this study attempts to capture students’ thoughts, feelings, and perspectives with regard to standardized testing through whole class discussions and individual journal responses. It will examine whether or not engaging students in practice tests and promoting reflection and discussion around test items can improve test performance. Finally, six focus students will be selected to participate in think alouds during testing and these same students will be interviewed at the end of the intervention to explain what aspects of the intervention each felt was most and least helpful and why.

**STATEMENT OF THE PROBLEM**

Historically, test preparation interventions give students insight about what a test will look like and what material will be covered on it, but such test preparation usually does not include in-depth analysis of specific test questions (Becker, 1990; Kulik, Kulik, & Bangert, 1984; Samson, 1985), especially, student-generated analysis. In the words of Kitsantas (2002), “Rather than simply posting or announcing the test grades in class, instructors should devote time to reviewing all test items following dissemination of test results, thus encouraging students to self-evaluate their test performance” (p. 111). Clearly, there is a need for in-depth analysis of test items to aid students in becoming more reflective. And, as noted by Zimmerman (2000), self-evaluation permits students to become aware of their deficiencies so that they are able to engage in strategy refinement.

Kitsantas (2002) and Zimmerman (2000) seem to suggest that a constructivist approach to test-preparation might yield benefits for students, particularly for those who score poorly or those who do not innately self-evaluate their progress. Such an approach
encourages students to construct their knowledge through interacting with others (Piaget, 1973; Vygotsky, 1978). As noted by Brooks and Brooks (1999), students in constructivist classrooms are encouraged to question each other, as well as to explain their thinking processes. Few test preparation programs elicit reflective writing in which students look back and evaluate their progress over several months. Thus, a constructivist approach to test preparation may be long overdue.

With the advent of the federal law, *No Child Left Behind Act* (2001), standardized assessments are mandated in every public school across America. Since the federal government is now monitoring the success of schools all over the country, the way schools operate is radically changing yet there are still stark achievement gaps among those socioeconomically disadvantaged. More than twice as many poor 8th graders perform at below basic levels in math than do those above the poverty line (Haycock, 2006). When reading levels and ethnicity are taken into account, the picture becomes bleak for everyone involved. Among 8th graders nationally, 49% of African Americans read at below basic levels, 45% of Latinos, 39% of Native Americans, 21% of Asians, and 19% of Caucasians (Haycock, 2006).

The trend of children of color performing poorly is also apparent on the CAHSEE, an exit exam designed to ensure that all students have a minimum level of skill in English and mathematics prior to high school graduation. Although current 2008 data shows that 78.8% passed English and 78.3% passed math on their first try (CDE, 2008), Hispanic and African American students fall well below these statewide averages. In addition, English learners and special education students experience pass rates under 50%. And not only are these special
populations scoring low on standardized tests like the CAHSEE, they are also quitting school more frequently.

In the state of California, one in four students quit school last year for an overall dropout rate of 24%. Among this population, 42% of African Americans along with 30% of Latinos dropped out, far more than the state average, and far more than state officials had thought ("California High School Dropout Rate," 2008). According to Jack O'Connell, California's Superintendent of Public Schools, "This is a crisis" (as cited in "California High School Dropout Rate," 2008). Clearly, high schools must do more to ensure that students pass this exam since a student's decision to remain in school may depend on it.

Students, particularly those at risk for dropping out, not only need quality subject-area instruction to pass an exam like the CAHSEE, they need teachers who provide test preparation systematically and thoughtfully, not in a disjointed, unorganized manner some few weeks before a state test (Chapman, 2002; Crocker, 2005). As Alfie Kohn (2000) contends, our current standardized testing system actually creates a double whammy for low socioeconomic students since such test preparation not only takes up a lot of time but actually encourages teachers to use a drill-and-kill instructional approach. "The use of high-stakes strategy only underscores the preoccupation with these tests and, as a result, accelerates a reliance on direct instruction techniques and endless practice tests" (Kohn, 2000, p. 37).

Quantitative meta-analyses on test preparation (Becker, 1990; Kulik et al., 1984) reveal that content instruction, test-wiseness, and test anxiety play important parts in determining the success of test-takers. Research also indicates that self-regulation processes aid test-takers (Kitsantas, 2002) and that high scoring test-takers use more self-regulation processes (Hong et al., 2006; McClain, 1983). It remains unclear, however, whether or not
students neglect to use effective test-taking strategies because they are unaware of them or simply because they are not motivated to utilize them. Pressley, Yokio, Van Meter, Van Etten, and Freebern (1997) showed that although many underachievers are aware of various reading strategies, they are not motivated to use such strategies. Do such students become motivated to use effective reading strategies when they need to pass an exit exam in order to graduate from high school?

In traditional high schools, test preparation is still seen as something that teachers have to do in addition to their curriculum, and it tends to occur some few weeks before a high-stakes test (Crocker, 2005; Durbin, 2002), far too late for many students who need to pass such minimum competency tests to graduate from high school. For greatest effectiveness, development of good test-taking skills should be a year-long activity incorporated into classroom instruction not a frenzied activity incorporated just a few weeks before a major assessment (Crocker, 2005). Crocker (2005) suggests 20 strategies for the teaching of effective test strategies including helping students to learn how to determine what a question is really asking, encouraging students to explain how they arrived at correct answers, and having students practice systemic strategies for choosing one best answer for multiple-choice items. These strategies have been discovered by a number of scholars such as Brown (1982), Chittooran and Miles (2001), Milman and Pauk (1969), and Smith (1991). Interestingly, none of the strategies named requires that students think aloud and/or write about their thinking with regard to standardized test-taking. Two aspects of the present study, the think alouds created by focus students and the overall journal reflections, seek to address these important areas.
PURPOSE OF THE STUDY

Although some researchers believe that standardized test scores are a necessary component to assess the quality of schools (Crocker, 2005; Phelps, 2005; Plake, 2005), and others feel that such tests are misleading and even detrimental to students within such school systems (Craig, 2004; Darling-Hammond, 2006; Kohn, 2000; Sacks, 1999), the purpose of this study is to see if a specific intervention can aid students in preparing for a high-stakes standardized test like the CAHSEE. The proposed intervention involves students taking several, miniature Princeton Review practice CAHSEE tests. It then requires that students explain their metacognitive approach (metacomprehension, problem-solving) to specific test questions during in-class discussions. Finally, it requires students to write about (critical thinking) the impact that in-class discussions may have on their test-taking ability, particularly how their approach to specific questions might change in the future. In addition, six focus students will think aloud during the test taking process and participate in interviews at the end of the study in which they explain which aspects of the intervention were least and most effective and why.

The intervention also explores if specific groups (advanced vs. basic, basic vs. below basic, advanced vs. below basic) improve at different rates over a 3-month period within a semester. Since male students with low skills are most at-risk for failing the CAHSEE and for dropping out of high school (CDE, 2008), it’s crucial to investigate the effects that a specific intervention may have. The purpose of the focus students is to examine six students’ test-taking processes closely. Half of these focus students will be historically strong test-takers and half weak. The final phase of the study will include an interview with each of the
focus students to investigate their perceptions about what aspects of the intervention were the most and least helpful.

While advanced students have a good sense of *feelings of knowing* (Hart, 1965, cited in National Research Council [NRC], 2001; Nelson & Narens, 1990) with respect to their learning since they can often predict, rather accurately, whether or not they know specific testing material, underachievers do not have the same clarity about what they know or do not know (Hong et al., 2006; Kim & Goetz, 1993; McClain, 1983). Although female students are beginning to outperform male students on large-scale assessments like the CAHSEE, there remain many questions about what types of test-taking skills these females possess. Little research has been done to explore how struggling students, or students with low skills, navigate the testing landscape. Although a substantial amount of research exists explaining how less confident students set lower goals and are less motivated (Locke & Latham, 1990, 1994, cited in Crocker, 2005; Zimmerman & Bandura, 1994, cited in Crocker, 2005), choose easier tasks (Sexton & Tuckman, 1991, cited in Crocker, 2005), exert less effort and persist for shorter amount of time on tasks (Bandura & Cervone, 1983, 1986, cited in Crocker, 2005; Zimmerman, 1995), there is sparse information regarding how these students use effective test strategies during standardized tests. However, it is vital to study students’ processes, even if under a test simulation, in order to identify ways to help them become more successful. It is also equally important to focus exclusively on particular students so they can convey what types of instruction help them to perform more effectively on standardized tests.

Currently, there is a lack of research investigating how students think about and make sense of actual test questions. In addition, SAT studies have primarily consisted of a college-bound population, students unlikely to fail a minimum competency test like the
CAHSEE. The literature attests that, to date, very few studies provide insight about students' experiences with standardized tests. However, Debard and Kubow's (2002) study of over 2,000 K-12 students revealed that students didn't feel that standardized tests measured how well their teachers taught and they expressed that such tests contributed to low morale among students. One aspect of this study, the focus student interviews, will explore why students hold certain perspectives regarding standardized tests.

**Research Questions**

The research questions driving this study are as follows:

1a. How, or in what ways, does a series of practice tests influence students' readiness for the actual CAHSEE exam?

1b. How do student groups perform on average relative to each other (advanced vs. basic, basic vs. below basic, advanced vs. below basic)?

1c. How does practice test achievement between more advanced and less advanced focus students compare?

2a. How, or in what ways, do a series of classroom discussions influence students' readiness for the actual CAHSEE?

2b. Specifically, what verbal articulation and efficacy patterns can be observed in the focus students?

3a. How, or in what ways, do a series of journal reflections influence students' readiness for the actual CAHSEE?

3b. Specifically, what written articulation and efficacy patterns can be observed in the student groups (advanced, basic, below basic)?

4a. How, or in what ways, do a series of think alouds influence students' readiness for the actual CAHSEE?

4b. Which aspects of the intervention do the focus students perceive to be most and least effective and why?
THEORETICAL FRAMEWORK

The theoretical lens used in this study is that of a constructivist, or social theory. Constructivists believe that students actively construct their knowledge through interaction with each other and the world around them (Piaget, 1973; Vygotsky, 1978). In constructivist classrooms, as noted by Brooks and Brooks (1999), teachers present curriculum from whole to part with an emphasis on big concepts; they listen to, value, and use student questions to guide the curriculum. In addition, they allow students to use real world raw data that has meaning.

The role of the teacher is to enhance cognitive development that is a collaborative effort between the learner and more knowledgeable party. In order to understand an individual's cognitive development, the individual must be studied within his/her specific patterns of social interaction. Therefore, a classroom environment in which students talk about their thinking would seem to be an integral part of understanding how students learn with respect to standardized tests. Characteristically, the teacher allows students to learn from one another. Student-centered teaching approaches include students working together toward accomplishing a goal, participating in discussions, inquiring, forming hypotheses, and gathering data to solve problems. The research of Polkinghorne (1988) supports the construction and reconstruction of shared social narrative histories of places such as schools in order to unearth stories that individuals think, feel and experience. The classroom teacher will be the main researcher and a co-constructed narrative between teacher researcher and students will result in action research (Maxwell, 1996). Such research has the ability to capture the essence of the classroom in its most authentic form. Such research has the power to improve a teacher's practice and positively impact future instruction for many.
Advantages of constructivism include the development of high-order thinking skills and problem solving (Jones et al., 2003). Although many researchers question the effectiveness of high-stakes testing (Barksdale-Ladd & Thomas, 2000; Wideen, O’Shea, Pye, & Ivany, 1997), claiming that it has led teachers away from inquiry-based learning, perhaps incorporating a constructivist approach on specific standardized test questions may lead to student improvement. Might giving students the opportunity to share their thinking around specific test questions deepen their understanding of how to be more effective on standardized tests? Might having students independently write about such discussions lead to changes in their awareness of the strategies they are using or are not using?

**SIGNIFICANCE OF THE STUDY**

The problem of low achievement on the CAHSEE is significant because it is a high-stakes test in which failure will jeopardize a student’s opportunity to graduate from high school. This study is important because educators need to know if certain intervention practices influence student test performance. In addition, this study strives to compare advanced, basic, and below basic students to investigate if specific behaviors are evident in students during, and upon reflection of the testing process. A final aim of this study is to investigate what is effective and not so effective with regard to test preparation from the students’ point of view since there is almost no literature on this topic. Since high failure rates for the CAHSEE and other standardized tests tend to be with male students with low skills (CDE, 2008), it is crucial to investigate why.

Finally, this study strives to provide a potential intervention model for a unique form of test preparation. In addition, it may also provide feedback to policy makers and
test-makers from a student perspective. Students may identify what types of test questions they collectively struggle with as well as identify areas of ambiguity on certain test questions. In an ideal system, such student feedback could be generated back to the parties responsible for test construction, thereby giving them the opportunity to build an even better assessment system. Perhaps such feedback can be a step toward improving our current testing culture.

**LIMITATIONS OF THE STUDY**

There are numerous limitations apparent in this study. The first is that the researcher responsible for conducting this study is actually the students' English language arts teacher. Although students will not be graded on the standardized testing intervention, because such an intervention occurs periodically within the normal curriculum, students may feel pressure to perform, believing, even erroneously, that the teacher researcher will withhold privileges if they choose not to participate.

In addition, the actual English Language Arts (ELA) CAHSEE consists of approximately 80 multiple-choice questions. While the number of test items in this study is comparable at 75, the number of items on each test is limited to 25. Therefore, students will be taking practice tests that are significantly shorter than the actual CAHSEE. Also, reading excerpts with more than seven or fewer than two questions have been omitted (there weren’t enough to match into three separate tests), grammar and related excerpts not connected to a text have been omitted (the study focuses on students’ ability to navigate text not grammar), and excerpts containing poetry have been omitted (there weren’t enough of these to place into three separate tests). In addition, no written responses were conducted for this analysis, a realistic section of the ELA CAHSEE. Therefore, no conclusions can accurately be drawn in
terms of predicting how students will perform on the actual CAHSEE. However, a goal of this study is to help students to improve their text navigation skills that may, in turn, increase their performance on the ELA CAHSEE.

In addition to such limitations, the actual physical limitations of the groups studied are apparent. Because of the superintendent’s mandate to make all freshmen classes no more than 20 students for this school year, the sheer number of students involved will not be typical of most freshmen classrooms across America. Because of smaller numbers, the teacher researcher may experience more success with a standardized testing intervention than another teacher who may have twice as many students. In addition, the sample populations studied will consist of one advanced group, a basic group, and a below basic group with few English language learners (ELL) and special education (SPED) students.

One of the criticisms of this study may be that it is not reflective of the student populations actually failing the CAHSEE, which are largely English language learners and special education students. To this criticism, the teacher researcher can attest that while the student population may not contain many ELLs or SPED students, it will contain a certain number of low achievers who may also be in danger of dropping out of high school. Another limitation is that only six focus students will be selected for in-depth analysis. These students, a mixture of males and females, and half high and half low test-takers, will not necessarily be representative of the sample population studied nor may they be representative of the entire population of students taking the CAHSEE.

Finally, the school environment in which this study will be conducted is also atypical, consisting of a 4 X 4 schedule in which students attend only four classes each day for approximately 85 minutes each instead of the typical 55-minute per class schedule of six
courses found in many high schools. This schedule, part of the small schools movement, a reform of comprehensive low-income high schools that began in the 1990s, allows for fewer courses and more personalization with teachers (Darling-Hammond, 2004), important factors in keeping students connected to school and making them less likely to drop out.

**DEFINITION OF KEY TERMS**

Since this study contains terminology that may be interpreted in multiple ways, for the purposes of this study, the following definitions will apply.

**Coaching**—A wide variety of test preparation activities undertaken by individuals in an attempt to improve test scores. Instructions given in preparation for taking a test that are designed to elicit maximum performance by the coached examinee (Cole, 1982).

**Concurrence seeking**—When members of a group emphasize agreement, inhibit discussion to avoid any disagreement or argument, and avoid realistic appraisal of alternative ideas and courses of action (Johnson & Johnson, 2007).

**Constructive controversy**—When one person’s ideas, information, conclusions, theories, and opinions are incompatible with those of another and the two seek to reach an agreement (Johnson & Johnson, 2007).

**Criterion-referenced test**—An assessment designed to sample specified curricular and/or instructional objectives on which an individual’s score is interpreted in relation to a cut score or standard (Crocker, 2005).

**Debate**—When two or more individuals argue positions that are incompatible and a judge declares a winner on the basis of who presented his or her position the best (Johnson & Johnson, 2007).
**Identical tests**—Test that appear exactly alike in nature. For example, some appear as pre- and posttests. It is thought that for test results to be reliable, there should be a year lapse between distributions of tests (Kulik et al., 1984).

**Metacognition**—Thinking about one’s thinking through a process of reflecting and analyzing, drawing conclusions from analysis, and applying such information to future thought practices (Downing, Ho, Shin, Vrijmoed, & Wong, 2007).

**Norm-referenced test**—An assessment on which an individual’s score is interpreted through comparison to the score distribution of a well-defined norm group who took the test at a given time (Crocker, 2005).

**Parallel tests**—Tests that appear different in nature but those that still assess the same content. It is recommended by Vernon (1954) to use this form of test over identical forms to create more reliable results (Kulik et al., 1984).

**Self-monitoring**—Deliberate attention to an aspect of one’s behavior (Kanfer, 1971).

**Self-regulation**—Self-generated thoughts, feelings, and actions for attaining goals (Zimmerman, 2000).

**Standards**—Descriptors that define the knowledge and skills that students should acquire at each grade level in order to accomplish the highest level of achievement (CDE, 2008).

**Test wiseness**—A subject’s capacity to utilize the characteristics and formats of the test and/or test-taking situation to receive a high score. Test-wiseness is largely independent of the examinee’s knowledge of the subject matter for which the items are supposedly measures (Milman, Bishop, & Ebel, 1965).
CHAPTER 2

REVIEW OF LITERATURE

This chapter will review research from five primary areas, each which will impact the study of freshman students in their quest to be successful on the CAHSEE. First, effects of the most recent federal law, No Child Left Behind Act (2001), will be explored followed by a brief history of standardized testing since its inception in America. This will provide a framework for understanding the value of standardized tests in U.S. culture, and how such tests have largely been mandated through policy. Next, studies about standardized testing will be explored, including examples of what contemporary schools have done to raise test scores. After that, research on effective ways to coach students on standardized tests will be discussed. Finally, an explanation of metacognition and discussion, along with their role in this study, will conclude the review.

THE EFFECTS OF THE No CHILD LEFT BEHIND ACT (NCLB)

Where once the teacher was the sole person who decided the curriculum, high-stakes testing puts the control firmly in the hands of politicians. This shift in power, respect, and professional autonomy leaves teachers caught between their beliefs in best practices and their need to show those in power what their students know and can do . . . . For the teacher in the classroom, who must face the public scrutiny of test scores, altering the curriculum is one way of gaining some measure of control over externally mandated testing policies for which he or she has little control. (Jones et al., 2003, p. 34)

The words of Jones et al. (2003) attest to the complex impact that the federal No Child Left Behind Act (2001) has had on educators throughout America.
Because high-stakes standardized tests have become the main source of measurement in American schools, the content of what is tested and, therefore, what is learned at school, has been profoundly affected. In the Jones et al. (2003) quotation, the authors acknowledge a shift in power away from the teacher and into the hands of politicians. In addition, they hint that in order for teachers to regain their autonomy and respect, they may have to sacrifice their best practices in favor of “teaching to the test.” Clearly, for many, making such a choice would be incompatible; for others, it might even be inconceivable.

Linda McNeil, whose research focused on the Texas accountability and assessment system (McNeil, 2000), concluded that teaching to the test narrows the possibilities of the curriculum greatly and puts extraneous limitations on what both teachers and students can achieve. Johnson, Finn, and Lewis (2005) discovered similar results in their research. Amrein and Berliner (2003) found evidence that standardized testing shakes students’ self-confidence while Debard and Kubow (2002), in their Ohio survey of over 2,000 students, discovered that standardized testing had a negative impact on student morale.

While the goals of NCLB are to improve the academic performance of all students, to achieve equity in and among schools, and to provide more qualified teachers (Darling-Hammond, 2004), its complex regulations for showing adequate yearly progress (AYP) toward test targets of 100% proficiency have created, what Darling-Hammond (2006) describes as “a bizarre situation in which most of the nation’s public schools will be deemed failing within the next few years, even many that already have high scores or that are steadily improving each year” (p. 654). Such multiple-choice and short answer exams result in less emphasis on complex thinking, communication and problem-solving skills that students need in order to be successful in the 21st century (Perkins-Gough, 2005). Although professional
testing experts warn against relying on any one measure as the sole source of important educational decisions (Berlinger & Biddle, 1995; Bracey, 1997; Gulek, 2003), currently California, and approximately a dozen other states, use single tests in order to issue high school diplomas (Perkins-Gough, 2005).

Both Darling-Hammond (2006) and Allensworth (2005) maintain that states like California, that require students to pass a single exam in order to graduate, may undermine low-achieving students the most, those who often have less familial support and lower educational skills. And finally, both Darling-Hammond (2006) and Boaz (1991) declare that our accountability system should measure individual student progress and not entire schools since the complexity and uniqueness of each school’s demographics naturally create an unequal playing field when subgroups are compared.

Yet while several researchers maintain that such testing is unnecessary and may even be harmful to students, many others believe that a testing system in which schools and students are held accountable is long overdue. Plake (2005) supports high stakes tests because such tests certify whether or not an examinee has the requisite skills and competencies needed to succeed in the workforce. In addition, such tests certify to the public that students are learning essential skills. As Phelps (2005) demonstrates, public support for standardized assessments has not wavered over many decades. Finally, Chudowsky and Pelligrino (2003) support large-scale standardized assessments so long as they reflect purposeful constructs and Scheurich, Skrla, and Johnson (2000) believe that, regardless of the problems inherent in our standardized system, some important truths have surfaced through their inception, such as our lack of academic success in educating children of color.
Before acknowledging how, specifically, certain school districts have handled the new challenges of accountability brought forth by NCLB, it is first imperative that a brief history of standardized testing in America is presented. While Scheurich et al. (2000) and countless others remark that our current educational system has failed to educate children of color (CampbellJones & CampbellJones, 2002; Causey-Bush, 2005) as well as immigrant children (Gandara, Rumberger, Maxwell-Jolly, & Callahan, 2003; Garcia, 2003), in looking back to the beginnings of standardized testing, it is clear that this practice has occurred ever since organized testing in the United States began.

THE EVOLUTION OF STANDARDIZED TESTING IN AMERICA

In 1837, when Horace Mann became the secretary to the first state board of education in Massachusetts, he helped to professionalize the field of teaching (Jones et al., 2003) and he was given the great responsibility of distributing one of the first American standardized tests (Sacks, 1999). As the Massachusetts school system grew, Mann faced pressure from the state school superintendent to improve and reform the schools. He responded by administering a test to measure individual achievement. Many of the test questions were tricky and early test-takers only got about 30% of the questions correct (Sacks, 1999). In “Mann’s Massachusetts,” as Sacks (1999) points out, “the availability of standardized test results compelled public officials to numerically compare and rank schools, even though the test’s purpose was to assess the achievement of individual students” (p. 71). Thus, these tests were misused shortly after they were developed.

In the late 1870s, the New York State Regents Testing Program, concerned with the educational level of high school graduates, constructed examinations to create a suitable
standard for secondary school graduation (Jaeger, 1982). This was the first hint of minimum competency tests, tests that would later drive the American assessment system. At the turn of the century, American schools began to group students by grade levels and according to age (Jones et al., 2003) and as the mass immigration influx expanded, the practice of using standardized test results inappropriately to compare children and schools, without considering influences not related to schooling, continued (Cremin, 1964).

Throughout the next few decades, associations between intelligence and heredity were made and went largely unchallenged until a seminal study began to question this concept. Might the test be the problem instead of the actual test-taker? In 1951, Eells, Davis, Havighurst, Herrick, and Tyler (as cited in Sacks, 1999) were the first to focus on the possibility that some measured differences in test scores were not accurate reflections of ability. This landmark study on test bias, which began as an Eells University of Chicago dissertation, shed this light by examining individuals’ responses to specific test questions, using group differences in percentage correct to analyze items (Camilli & Shepard, 1994). Their work helped society to at least begin to question whether ability was in fact, determined by heredity.

As America moved into the second half of the 20th century, the federal government, for the first time, created policy that gave supplemental funding to its states. The results of this funding over time, has resulted in greater and greater control. Beginning in the mid 1960s, the Elementary and Secondary Education Act (ESEA) of 1965 was created to provide revenue for textbooks and other instructional materials. The primary purpose of the act was to provide children from low-income families with adequate resources. Because no school subject was considered to be more important than reading, a substantial portion of the
funding of ESEA was earmarked for remedial programs in reading under the Title I provision of the act (Ruth, 2001), a program that is still largely funded today in low socioeconomic schools. Coincidently, where standardized tests once served primarily as indicators of achievement from the 1920s to the 1960s, they now began to become the primary instrument of accountability as state legislators began to use such tests as mechanisms of power to leverage reforms (Ruth, 2001).

Although low socioeconomic schools benefited from receiving Title I funds, the money came with strings attached. The federal government began to monitor schools using the results of norm-referenced standardized tests, such as the Iowa Test of Basic Skills (Sacks, 1999). By requiring that Title I schools be evaluated by means of test scores, the law effectively mandated states to employ standardized tests in order to receive funding. ESEA’s effect on expanding standardized testing in American schools became, in the words of Sacks (1999), “unquantifiable” (p. 75).

And apparently, according to the 1983 release of A Nation at Risk (National Commission on Excellence in Education, 1983), states weren’t progressing very well. The report compared the nation’s children with those of other countries and according to all outward appearances, America stacked up close to the bottom. Perhaps no other political report can compare to the impact that its release had on the general public. In the words of Ruth (2001), “it jolted the nation with its grim message about ‘the rising tide of mediocrity’ in our educational system” (p. 97). The report set off a spiral of subsequent reforms including a demand for back to basics rigor, systemic reform of textbooks, curriculum, testing, and teacher preparation requirements, and even private management of schools (Ruth, 2001). It also contributed to fueling what the federal government began in 1994, the Goals 2000:
Educate America Act. Calling for improvements, the act jumpstarted the launching of state standards. By 1996, 14 states had developed them. By 1999, all with the exception of Iowa had detailed academic standards (Ruth, 2001).

The federal government, through Goals 2000, was rearing its head again, and this time, the public officials were truly taking note. In Chicago, 1 in 10 of their 424,000 students were sent to mandatory summer school on the basis of standardized test scores (Sacks, 1999). In Milwaukee, Wisconsin, the school district agreed to pay over $400,000 for a computerized program called TargetTeach. The firm that supplied the product promised a 20% to 200% growth in test scores in a year (Sacks, 1999). In California, the performance-based CLAS assessment, an innovative test attempting to replace a previous standardized test, was vetoed by Governor Pete Wilson (Ruth 2001; Sacks, 1999). In its place, the state legislature developed STAR, the Standardized Testing and Reporting program, requiring all children in grades 2-11 to take criterion-referenced, multiple-choice tests modeled after the Stanford Achievement Test (Ruth, 2001). Throughout the country, an entire slew of high school exit exams, a return to minimum competency tests from the late 1970s and early 1980s, would soon follow (Condiffe-Lagemann, 2000; Cuban, 1997; McQuillan, 1998).

THE EFFECTS OF STANDARDIZATION

While many, particularly the naïve public, have assumed that test score improvements within high-stakes assessment systems equate to a superior education, others, within such systems, have found the opposite may be true. In her North Carolina study of 13 middle school language arts teachers, Watanabe (2007) concluded that, in all five schools observed, teachers felt that testing and test preparation took time away from teaching and narrowed the
scope of writing genres. As Watanabe (2007) notes, “The essentialist testing policy appears to encourage teaching that runs contrary to the goals that the state wants to reach” (p. 354).

Watanabe (2007) concluded that there are grave problems in the present accountability system in North Carolina because the standards and the assessments are at odds with each other. While the North Carolina state standards suggest students respond to standardized tests with right answers, the state standards encourage students to become “complex thinkers who possess the ability to reason, make decisions, and solve problems in a variety of contexts” (Watanabe, 2007, p. 340). As Watanabe (2007) points out, while students explored practice test questions, “The focus on right answers did not afford students the opportunity to construct meaning from the text” (p. 341).

In Scherff and Piazza’s (2005) descriptive quantitative survey study analyzing over 2,000 Florida students’ perceptions of how they wrote and the extent to which they wrote in their English language arts high school classrooms, data revealed that although writing instruction occurred in a variety of ways and in different capacities, it was often at odds with research-based practices. The researchers speculated that such an effect may be the result of existing in a harsh political climate, a place in which school accountability and statewide assessments truly are influencing the way that teachers teach.

In Jeffrey Durbin’s (2002) study of Jackson High School in Chicago, he noted the trend for high amounts of test preparation prior to a high-stakes exam. “During the school year, students did various kinds of test preparation activities but the real push began about two months before the TAP [state standardized test] was given” (Durbin, 2002, p. 70). In Durbin’s interviews of students, he reports that students tired easily and were bored with repeating the same kinds of fragmented problems and activities and that the majority of
students felt that test preparation wasted their time. Several complained that they could have been learning something worthwhile instead. One student described his test preparation experience as “doing a little packet” (Durbin, 2002, p. 74).

Not only do standardized assessment systems appear to influence the way teachers teach, they also appear to adversely affect minority children in particular. In their quantitative analysis of Florida students, Borg, Plumlee, and Stranahan (2007) examined the Florida Comprehensive Achievement Test (FCAT) results for all 10th grade students in Duval County, Florida, during the years of 1999-2000 ($N = 5,206$) to address the question, has the enforcement of standardized testing uniformly benefited children of Duval County or are there uneven and troubling effects? Borg et al. made several findings including that the aggregate, average race sample was 54%. However, when various races were disaggregated, the results changed dramatically (white = .65, African American = .34, Hispanic = .54). In addition, when socioeconomic status was factored in, higher income student pass rates were .60 while low-income pass rates were only .30. A significance of $p < .01$ was established in white vs. African American student characteristics (.31) as well as high income vs. low income (.30).

Borg et al. (2007) concluded that minority students with higher mobility rates coming from poorer, less educated backgrounds are less likely to meet graduation requirements than are students coming from white, suburban, wealthier, and more educated homes. And finally, instead of a standardized assessment system closing the achievement gap, it appears that the performance testing approach might actually worsen the gap between African American and Hispanic students and their white counterparts. The researchers concluded that, “The social difficulties imposed by race and class differentiations in this country will virtually ensure that
schools with large percentages of poor and minority students will be more likely to have lower average student scores on standardized tests such as the Florida’s FCAT” (Borg et al., 2007, p. 705).

While comparisons between ethnic student groups may account for some of the high variance in test scores, perhaps the quality of the standards in each state and the tests used can account for the other part. Through NCLB, states can arbitrarily set their own standards, making some appear to outperform others when, in fact, inferior standards and/or tests may actually be the case. In her analysis of four southwestern states, Paula Garcia (2003) evaluated the English language learner (ELL) passage rate on high school exit exams in four states (New Mexico, Texas, Arizona, California) between the years 1999 and 2003. Garcia discovered that in all four states, a high school examination was required in order to graduate; however, some states gave special accommodations to ELLs while other states did not. In the states of New Mexico and Arizona, accommodations were made for ELLs. Such accommodations included oral reading of test directions, clarification of English words, and the use of bilingual dictionaries. In addition, the test was also administered in small groups, often in the students’ bilingual classroom. In California, the only provision given was extended time, a provision also granted in the state of New Mexico. In Texas, no accommodations were given even though this state has a large ELL population.

And not only are such standard discrepancies affecting English language learners, they are also affecting African American children on a grand scale. In Tonia Causey-Bush’s (2005) descriptive study comparing 2003 results from the California Standards Test (CST) with the Texas Assessment of Knowledge and Skills (TAKS), she found that when comparing overall test results on the CST and the TAKS, the numbers are staggeringly
different. In English language arts, 84.5% of students in Texas scored proficient or above while only 36.5% of California's students did. In Texas, 77% of African Americans scored proficient or above while only 23.8% did so in California; 79% of Texas Hispanic students scored proficient or above while only 21% did so in California. In Texas, 91.7% of white students scored proficient or better, while only 55.1% did so in California. These same dramatic gaps could be witnessed in the ELL, SPED, and socioeconomically disadvantaged populations, with Texas trumping California every time in every category.

Causey-Bush (2005) concluded that the difference in student achievement between California and Texas may likely have to do with "differences in curricular coverage than they do in the difficulty level of the TAKS or the CSTs" (p. 340). In other words, Texas may do a much better job of teaching to the test. Because California and Texas collectively educate 34% of the nation's children of color (N = 6,432,254), Causey-Bush called for an immediate National Assessment for Educational Progress (NAEP) test comparison between California and Texas to see how each state's students would measure up using the same test.

Regardless of whether certain states have an effective standardized testing system in place is a topic that can be debated for as long as such systems remain; and if the history of testing in America is any kind of predictor, such systems will be here for many years to come. The more critical and immediate concern, however, is how do educators prepare students to be successful on such tests? While the preponderance of evidence suggests that many teachers feel that test preparation takes up too much class time, often perceiving it as an added burden that detracts from their curriculum (Durbin, 2002; Herman & Golan, 1990, cited in Sacks, 1999; Jones et al., 2003; Shepard & Dougherty, 1991), there is little research to document what types of test preparation are being offered to students, and whether or not
such test preparation is effective, regardless of the time it may take to institute. The next section gives several examples of types of test preparation that have been shown to be effective while the final section takes into account the approaches needed with different learners and justifies the use of a metacognitive (metacomprension, problem solving, critical thinking) approach when preparing students for high-stakes tests.

**Effective Test Preparation**

To date, research on teaching methods that help students pass the CAHSEE or any other standardized American graduation exit exam do not appear to exist. After an extensive review of online test preparation programs, it was discovered that Princeton Review and Kaplan don’t offer CAHSEE courses but instead focus on SAT/ACT test preparation type courses. Of the companies that do offer CAHSEE test preparation such as USA Test Prep., ALEKS, Test Prep. Review, XL Prep., and Study Island, all are offered only online; none offer a participatory face-to-face aspect. The current approach to test preparation appears to be solitary, ignoring the importance of community participation. In addition, among studies that have been conducted about test preparation, very few concretely describe the learning environment in which test preparation activities take place. This study strives to capture that environment through various “snapshots” of student thinking about practice test questions over time. It also strives to see if the differences between high and low achieving students described in the literature are, in fact, apparent. Finally, this study attempts to capture student thinking through recorded think alouds and written journal reflections about test questions to ascertain if these added components affect the learning experiences of specific focus students.
In Pedal Walker's (2007) dissertation that focused on two U.S. History Advanced Placement (AP) teachers who generated high pass rates among low-income students, she found that characteristics of successful AP students, even from low-income environments, tended to be self-reliance and resilience. While both teachers in the study embedded test preparation in their curriculum to some extent, they expected students to be “the main drivers of their own learning” (Walker, 2007, p. 193). They did not flexibly respond to the students’ personal academic needs. In both cases, students were responsible for handling the bulk of the work (accomplishing outside readings and essay drafts, etc.). In her conclusion, Walker recommends not to coddle low-income students but to challenge them as they develop “the durability to endure challenging work” (p. 196). Yet as Vygotsky (1978) attests in his famous Zone of Proximal Development model, if there is too much of a gap between the learner’s skill level and that of his/her more capable peers or mentors, the challenge to strive may not be accepted. The assumption, on the part of these AP teachers, seemed to be that all of the students were capable and motivated and, being in an advanced placement course, that may in fact, be correct. But what happens when a group of students isn’t so driven and doesn’t seem to have that self-reliance or resilience?

While research suggests that there are important processing differences between high and low achievers when they test (Kim & Goetz, 1993; McClain, 1983), the research is not so clear about what might account for these differences. In Walker’s (2007) dissertation, one of the exemplary teachers offered simulations on practice material similar to that on the actual test (Pressley et al., 1997), guided students in how to narrow down multiple-choice questions (McClain; 1983), and embedded test preparation throughout the content being discussed (Becker, 1990; Crocker, 2005). The other teacher tended to frontload his students with
multiple-choice practice exams throughout the year, but the researcher didn’t provide any discussion on how, specifically, he discussed these items with his students, only that a month before the actual test, the classroom activities consisted of “non-stop test preparation” (Walker, 2007, p. 88), a pattern seen in other high-stakes testing environments (Durbin, 2002; Shepard & Dougherty, 1991). Thus, even though both teachers were considered exemplar because of high student pass rates, there was still a bit of a mystery about what they actually did when they discussed test items with students. This mystery is apparent in many other studies and may limit the conclusions that can be made from larger meta-analyses like Kulik et al.’s (1984), Samson’s (1985), or Becker’s (1990).

Before discussing studies in which teachers explore students’ processes during multiple-choice test items, it is first necessary to clarify that the majority of test preparation studies, even those recently conducted, have focused on the Scholastic Aptitude Test (SAT), a popular, norm-referenced test that is required for admission into most competitive colleges. Constructed by Lewis Terman in 1922 originally as an IQ measure, it currently measures verbal, mathematical, and written ability of its subjects (Sacks, 1999). Although it has undergone a number of changes since its introduction, its primary purpose is still to predict how well students will perform academically in college (Eckland, 1982). It is not an achievement test that measures what has been learned. This is an important distinction.

Since the mandate of NCLB, the federal government currently measures states based on criterion-referenced tests that are designed from state standards. Therefore, a meta-analysis for other forms of tests may not be generalizable for criterion-referenced test studies. As Vernon (1954) recognized long ago, “The majority of these [40 relevant research reports], however, show serious technical difficulties” (p. 270). Vernon then names one of
the difficulties as studies that compare identical tests results (exact test versions) with parallel forms (differing versions of the same test). He also identifies more incongruities when comparing studies, including whether or not subjects learned solely from their own experience on tests (repetition of tests) or whether they were given hints on how to improve their performances (part of coaching).

Yet even though the conclusions drawn from meta-analyses are limited because the conditions of testing are so varied, some trends are apparent and add to the research base. In their meta-analysis of 40 SAT studies, Kulik et al. (1984) found that the effects of test gains were greater as the number of practice tests given increased. Thus, the more opportunities students had with practice tests, the better they appeared to do. Powers (1988, cited in Becker, 1990) found that, when giving practice tests, coached students benefited from a systematic review of the content domain more than those who were not coached. In addition, like Vernon (1954), Kulik et al. discovered that students improved more readily when identical forms of the test were used as opposed to parallel ones, and added the recommendation that because identical tests have the power to influence subsequent trials, they should not be used to accurately measure changes in students’ scores. Thus, the beginnings of a testing framework emerged. Thirdly, coaching appeared to have contributed to test score gains. Students coached 9 hours or more showed a greater effect size (.16) that those coached 3 to 9 hours (.08; Kulik et al., 1984). Messick and Jungeblut (1981) found similar results in their work, concluding that least 3 hours of coaching was necessary to show gains.

However, other studies show that coaching gains are minimal, at best. In her study of 23 SAT studies, Becker (1990) stated that for every 10 hours of coaching, a coached group
will only gain 1 to 6 points on the SAT. Seaton, Underwood, and Fortune (1992) found similar results when investigating the effects of ACT coaching on 30 junior girls and found students who received 10 hours of preparation increased their composite scores an average of 4 points. When one attempts to compare the coaching methods used in Kulik et al.'s (1984) or Becker's (1990) meta-analysis, the details about the type of instruction occurring during such test preparation are scant and inconclusive. Therefore, it is difficult to know what forms of coaching are effective.

While Becker (1990) reached many similar conclusions as Kulik et al. (1984), she added other noteworthy aspects such as the idea that familiarizing an examinee with novel item types may well enable him or her to improve SAT performance considerably. Becker stressed that content-relevant instruction in coaching groups showed not only increased scores, but increased advantages. She raised questions about fairness for low socioeconomic groups since certain groups paid for a test preparation course while other volunteered via school elective or after school program. Because low socioeconomic students would most likely find voluntary programs more accessible and desirable than paid programs, Becker raised the point that not having information about how coached groups were selected might severely compromise the validity of the data.

Other scholars like Mehrens and Kaminski (1989), Popham (1991), and Haladyna, Nolen, and Haas (1991) strongly warn against developing a curriculum based on a test or presenting items similar to those on a test because they consider it unethical. Rather, they feel that the content should be the main focus of the instruction, and that the tests should be designed to measure the content that students actually know. While the state standards movement in the late 1990s attempted to create standards that are linked with
criterion-referenced tests, this system has been criticized by testing experts because the content measured on such tests is not able to measure all of the standards equally. Some areas are weighted more than others so the validity is compromised when groups are compared.

In her analysis of Kulik et al.'s (1984) work, Becker (1990) acknowledged that programs that incorporated test-wiseness, test-anxiety, or content-instruction made the most gains. Although larger gains for high ability students than for low ability students have been shown (Kulik et al., 1984; Samson, 1985), these same studies don’t provide information about how high vs. low students are processing while they test. The speculation is that high ability students have a greater skill set or, in Vygotsky’s (1978) model, they are in the Zone of Proximal Development. Low ability students may not have this skill and may need skill development before they can advance at a higher rate (Beidel, Turner, & Taylor-Ferreira, 1999). A few studies explore the processes that high and low achievers experience when they test which is helpful in constructing an effective model for this study.

In her think aloud study of how “A,” “C,” and “F” students test (N = 60), McClain (1983) speculated that gains observed in her “A” level college students might be attributed to superior reading skills and better ability to verbalize their thinking. She came to this conclusion because the high “A” students in her study considerably outperformed the low “F” students and even the middle “C” students in virtually every way on her 70-question psychology course exam. The greatest differences could be seen in the “A” students’ ability to anticipate multiple-choice answers without looking at the a, b, c, and d choices. “A” students anticipated, on the average, 40.9 times out of 70 questions, whereas “C” students only did this on 7.65 questions, while “F” students did this on only 2.3 questions. In addition, “A” students read through an average of 3.66 answers out of 4 answer choices before
selecting an answer whereas "C" (1.82) and "F" (1.48) students were likely to read only half as many alternatives before making a selection. This suggests that "C" and "F" students stop looking at alternatives once they believe they have found the answer.

Not only did "A" students anticipate answers more frequently and consider all alternatives before selecting an answer, they also explained why they eliminated incorrect alternatives more often. In this study, students could give a maximum number of three reasons for eliminating per test question, one for each answer they eliminated. "A" students gave reasons for eliminating incorrect alternatives at an average rate of 1.93 whereas "C" students did so at rates of .20 and "F" students did so at rates of .07. "A" students gave reasons for eliminating nearly two of the three choices on each test question. "C" and "F" students, on average, did not even give reasons for eliminating any of the choices on each test question. Although students with "A" averages clearly outperformed "C" and "F" students, it's not entirely clear why. What rationale did students have for eliminating choices? Did the "C" and "F" students not give reasons as frequently as the "A" students because they didn't know the content or is it because they chose not to articulate?

While McClain (1983) shared her results with her students and noted that some of the "C" and "F" students used the strategies characteristic of "A" students on future exam scores and reported improvement, she did not review the effective strategies in great depth, something that this study strives to do. This study attempts to help students to become better test-takers by allowing them multiple opportunities to "practice" taking tests. In addition, students repetitively discuss their thinking processes in a classroom environment, at least six times. Finally, this study contains a journal component in which students capture their knowledge gleaned through classroom discussions in writing at least six times over a
3-month period within a semester. One of the ways to develop test-taking skills is to teach students to use self-regulatory strategies to enhance their test-preparation and test-taking skills, thus enhancing their academic success (Hong et al., 2006; Kitsantas, 2002).

While there are differences in the ways that high and low achievers approach tests, certain specific behaviors have been shown to increase test effectiveness. In her analysis of self-regulation of 62 college psychology students, Kitsantas (2002) found that high test scorers tended to seek information and help at nearly every opportunity whereas low achievers were four times less likely to seek help. In addition, high test scorers reflected at a rate six times more often than low test-takers. Thus, providing a classroom learning environment in which students can receive assistance on their test-taking processes seems key, in addition to providing ways that students can reflect on their processes. In addition to seeking help and reflecting more often, high test-takers also used the process of elimination strategy more often. They also reported more goal setting and planning and monitoring of their progress. Finally, upon receiving results, high achievers used self-evaluation to make judgments of their test performances and to improve these performances on future tests. Schunk (1989) has shown that self-evaluation of one’s learning processes enhances students’ self-efficacy and motivation to continue working on a task.

In Kitsantas’ (2002) study, not only did high self-regulated examinees score significantly better on the test and report greater self-efficacy beliefs, high self-regulated students perceived the test as more important than did low self-regulated examinees. Such learners are also more likely to replace their strategies with more effective ones (Zimmerman & Martinez-Pons, 1986). Self-efficacy beliefs play a key role in students’ selection and use of learning strategies and goals (Kitsantas, 2002). This may be because self-regulated learners
are more also more self-motivated (Zimmerman, 2000). Based on Kitsantas’ study, it would seem very important to provide students with opportunities to reflect on their test performance. In her words, “instructors should devote time to reviewing all test items following dissemination of test results, thus encouraging students to self-evaluate their performances” (Kinsantas, 2002, p. 111).

Although there is considerable attention devoted to standardized tests and particularly, high school exit exams, since failure of these tests can stop a student from graduating, there is little information as to what types of test preparation instruction are effective for students. While large meta-analyses have ascertained that content area instruction, test wiseness skills, and test anxiety reduction strategies are important contributors to assist students in becoming more successful on multiple-choice tests (Becker, 1990; Kulik et al., 1984), there is little information about what successful test preparation environments look like. However, developing students’ metacognitive skills through discussion can provide evidence of improved higher order thinking, which may, in turn, manifest improved results on a series of practice tests.

**METACOGNITION**

Although the term metacognition only became introduced into academic scholarship in the 1970s when Flavell (1971) coined the term metamemory, the concept of reflecting upon one’s thinking processes dates back to Socrates’ profound notion that “the unexamined life is not worth living” as recorded by Plato. Currently, The National Research Council (NRC) defines metacognition as “the process of reflecting on and directing one’s own
thinking” (2001, p. 4), while others define metacognition as the monitoring and control of thought (Martinez, 2006; Nelson & Narens, 1990).

Curiously, Flavell (1999) defines metacognition as simply “thinking about thinking.” Perhaps a more modern definition, reflective of the challenges in current schooling, would be Downing et al.’s (2007) definition that metacognition not only involves thinking about thinking, but how to reflect and analyze thought, how to draw conclusions from that analysis, and how to put such information into practice. Clearly, metacognition has become a construct with multiple steps and dimensions.

**Metacognition and the Classroom**

Classroom research has demonstrated that at least three specific techniques are especially useful in making thinking skills explicit: modeling, metacognitive reflection, and thinking aloud (Brown, Campione, & Day, 1981; Rosenshine & Meister, 1992; Sternberg, 1984; Taba, 1965). Modeling has been particularly helpful in introducing a new skill. This technique makes explicit and explains the steps in a thinking strategy or procedure while walking students through it step by step (Palinscar & Brown, 1984). The modeled skill procedure also provides a starting point from which novices can gradually develop more personalized procedures for carrying out a skill (Pressley & Harris, 1990).

Metacognitive reflection, when used to introduce a new thinking skill, can engage students in verbalizing, analyzing, and reflecting on what, step by step, they recall doing mentally to apply a just-completed thinking skill (Beyer, 2008). More importantly, repeated use of metacognitive reflection during initial efforts to apply a new skill enables novice students to identify flaws in their own thinking as well as recognize and gradually construct
or reconstruct more effective procedures for applying a skill (Brown et al., 1981; Nickerson, 1989; Nickerson, Perkins, & Smith, 1985; Paris & Winograd, 1990; Whimbey, 1980). As Vygotsky (1962) explained many decades ago, high-order thinking skills begin as social discourse and it is these discourse patterns that are internalized over time and through experience.

Thinking aloud has long been an important research tool for making thinking explicit (Beyer, 2008). It has been used to identify the thinking processes of novices as well as experts in writing, reading, comprehension, problem solving, and studying for tests. When a teacher “thinks aloud,” particularly during problem solving, his or her verbalizations can be a powerful source of cognitive processing that students can internalize (Martinez, 2006). Other researchers echo this concept.

Arthur Whimbey and Jack Lochhead (1999) developed a self-instructional program in which pairs of students think aloud while engaging in a problem-solving or analytical-reasoning task and then compare how they completed the task with an expert’s step-by-step approach. This process repeats several times for each type of task presented (Whimbey & Lochhead, 1999). The repetitive process appears to allow students of all levels to improve their processes. As Beyer (2008) notes, “Repeated use of metacognitive reflection during initial efforts to apply a new skill enables novices to identify flaws in their own thinking as well as recognize and gradually construct or reconstruct more effective procedures for applying the skill” (p. 226).

In addition to modeling, reflecting, and thinking aloud, building metacognitive knowledge of oneself as a learner contributes to viewing oneself as an able learner, which influences not only success in later learning, but also motivation to learn (Zimmerman,
1989). There is also evidence that metacognitive development, in addition to helping learners take charge of their own learning (Brown & Campione, 1996), contributes to feelings of self-efficacy in learning (Pintrich & De Groot, 1990) which, in turn, increases students’ motivation and interest in learning.

**Expert vs. Novice Metacognition**

Research in the past few decades has made clear distinctions between those with expert and novice metacognition. Experts have strong metacognitive strategies (Hatano, 1990). They monitor their problem solving, question limitations in their knowledge, and avoid overly simplistic interpretations of a problem. In the course of learning and problem solving, experts display certain kinds of regulatory performance, such as knowing when to apply a procedure or rule, predicting the correctness or outcomes of an action, planning ahead, and efficiently apportioning cognitive resources and time (Kitsantas, 2002).

This capability for self-regulation and self-instruction enables advanced learners to profit a great deal from work and practice by themselves and in group efforts. Strong learners can explain which strategies they used to solve a problem and why, while less competent students monitor their own thinking sporadically and ineffectually and offer incomplete explanations (Chi, Bassok, Lemis, Reiman, & Glaser, 1989; Chi & VanLehn, 1991). Studies of metacognition have shown that people who monitor their own understanding during the learning phase of an experiment show better recall performance when their memories are tested (Nelson, 1996). Thus, those students with a constructive learning conception may equate learning with seeking understanding (Purdie & Hattie, 2002) rather than merely seeking to memorize knowledge. Martinez (2006) divides the complex notions of
metacognition into three areas: metamemory and metacomprehension, problem solving, and critical thinking. These areas work especially well as a construct for the students taking, discussing, and then reflecting on multiple-choice reading comprehension questions on a variety of topics.

A Metacognitive Framework for Classroom Discussion

Some researchers claim that metacognitive ability is an important variable in predicting reading comprehension. Britton, Stimson, Stennent, and Gulgoz (1998) developed an individual differences model of learning from text in which they argued that making connections among ideas in text depends upon four variables. Britton et al. argue that students must first sense that a connections between concepts needs to be made. They must recognize a gap in their understanding. In order to bridge this gap, students rely on working memory, domain knowledge, and inference-making to understand text. Metacognition is the trigger for other processes that are necessary for understanding. These researchers argue that poor metacognitive ability interferes with these processes and that metacognition serves as a sort of prerequisite for comprehension.

In the area of testing, some researchers advocate using practice tests. Dunlosky, Rawson, and McDonald (2002) recommend using practice tests that are similar but not identical to class exams since students rarely have access to actual class exams. While practice tests often do improve performance both for learning relatively simple material and for learning more complex text, the role practice tests may play in bolstering metacomprehension accuracy is not well understood (Dunlosky et al., 2002). To benefit fully from practice tests, students may need instruction on how to interpret outcomes from such
tests. One alternative way is to make sure that students always have access to the correct
results on practice tests so that their own self-generated feedback can be supplemental with
perfectly accurate feedback (Dunlosky et al., 2002). Maki and McGuire (2002) suggest
investigating metacomprehension in a setting where more decisions about reading are made
necessary by the medium of text used.

In addition to providing practice tests and helping students to identify gaps in their
learning, problem solving ability is a crucial aspect of improving test results. Experts use
metacognitive strategies for monitoring understanding during problem solving and for
performing self-correction (NRC, 2001). Research has shown that self-insight has predictive
power for subsequent learning (Martinez, 2006). Metacognition is crucial to effective
thinking and problem solving and is one of the hallmarks of expertise in specific areas of
knowledge and skill. Practice and feedback are critical aspects of the development of skill
and expertise (NRC, 2001). Problem solving involves continuously generating possibilities,
weighing those options, exploring subsets of opinions, and evaluating results (NRC, 2001).

Like problem solving, critical thinking in an important aspect of the overall
metacognitive process. Critical thinking may involve evaluating ideas for their quality,
especially judging whether or not they make sense (Martinez, 2006). Students should be
encouraged to cultivate their metacognitive capacity. If students are encouraged and guided to
think critically together, then their spoken reasoning will ideally make their cognitive tools
available to one another (Martinez, 2006). Clearly, a pedagogical approach guiding students’
critical thinking and encouraging their spoken reasoning could be well realized in a
classroom setting positioned around the discussion of practice multiple-choice test questions.
Pressley et al. (1997) have argued that reading skills are not being addressed in secondary high schools in America. Can these skills be addressed through reviewing the material on practice English language arts tests since these tests measure students’ ability to read content presented on the test rather than content learned previously? Can students’ metacognitive processes be revealed when they review practice ELA Princeton Review CAHSEE test questions? More importantly, through discussions and reflective journal writing, can low achievers increase their capacity to behave more like high achievers when taking standardized tests? Since educators are currently, in the words of state superintendent, Jack O’Connell (as cited in “California High School Drop Rate,” 2008), experiencing a a “crisis” in the state of California due to our high dropout rate, educators must act quickly to help students, particularly African American and Hispanic male students with low skills, those most likely to fail the CAHSEE. Giving such students effective test-taking skills early in their high school years, and helping them to develop confidence to pass the CAHSEE, may in fact, give them reason to not only pass such an exam, but to stay in school and graduate, giving them more possibilities for their future.

**Constructive Controversy as a Means to Engage Students in Discussion**

An effective framework in which to measure both the quality and capacity of classroom discussions in this study will be through Johnson and Johnson’s (2007) constructive controversy model. According to Johnson and Johnson, constructive controversy is “when one person’s ideas, information, conclusions, theories, and opinions are incompatible with those of another and the two seek to reach an agreement” (2007, p. 38). Constructive controversy differs from concurrence seeking in that it generates
reconceptualization, synthesis, and integration resulting in higher achievement whereas concurrence seeking invites a quick compromise to a dominant view (Johnson & Johnson, 2009). Similarly, debate often involves a closed-minded adherence to one's own point of view (Johnson & Johnson, 2009), one in which external force decides winners and losers (see Table 1).

Table 1. Constructive Controversy, Debate, and Concurrence Seeking

<table>
<thead>
<tr>
<th>Process</th>
<th>Constructive Controversy</th>
<th>Debate</th>
<th>Concurrence Seeking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial conclusion</td>
<td>Categorizing and organizing information to derive conclusions</td>
<td>Categorizing and organizing information to derive conclusions</td>
<td>Categorizing and organizing information to derive conclusions</td>
</tr>
<tr>
<td>Oral presentation</td>
<td>Presenting, advocating, elaborating position and rationale</td>
<td>Presenting, advocating, elaborating position and rationale</td>
<td>Presenting, advocating, elaborating position and rationale</td>
</tr>
<tr>
<td>Level of uncertainty</td>
<td>Being challenged by opposing views results in conceptual conflict and uncertainty about correctness of own views</td>
<td>Being challenged by opposing views results in conceptual conflict and uncertainty about correctness of own views</td>
<td>Being challenged by opposing views results in conceptual conflict and uncertainty about correctness of own views</td>
</tr>
<tr>
<td>Motivation</td>
<td>Epistemic curiosity motivates active search for new information and perspectives</td>
<td>Close-minded rejection of opposing information and perspectives</td>
<td>Apprehension about differences and closed-minded adherence to own point of view</td>
</tr>
<tr>
<td>Revised conclusion</td>
<td>Reconceptualization, synthesis, integration</td>
<td>Close-minded adherence to own point of view</td>
<td>Quick compromise to dominant view</td>
</tr>
<tr>
<td>Relative outcomes</td>
<td>High achievement, relationship quality, and psychological health</td>
<td>Moderate achievement, relationship quality, and psychological health</td>
<td>Low achievement, relationship quality, and psychological health</td>
</tr>
</tbody>
</table>


Although some scholars feel that conflict is constructive in the classroom (Johnson & Johnson, 2007; Johnson, Johnson, & Johnson, 1976), others contend that it is undesirable,
and may lead to anger, hostility, rejection, damaged relationships, and decreased commitment to learning (Chiu & Khoo, 2003; Collins, 1970; Janz & Tjosvold, 1985). Surprisingly little conflict tends to be structured in instructional situations (DeCecco & Richards, 1974) and discussion remains a rarity in U.S. schools (Parker, 2006).

While it appears that there is some disagreement in the research community regarding conflict, the context of conflict may be an important indicator as to whether it is effective or not effective in the classroom setting. According to Johnson and Johnson (2003), two possible contexts for conflict exist, cooperation and competition. In the Johnsons' understanding, a cooperative context tends to facilitate constructive controversy, whereas a competitive context tends to promote destructive controversy.

In a cooperative context, constructive controversy tends to result in open-minded inquiry that leads to refined conclusions (Johnson & Johnson, 2007). In addition, constructive controversy in a cooperative context tends to induce feelings of comfort, pleasure, and helpfulness in discussing opposing positions, an open-minded listening to the opposing positions in addition to motivation to actually hear more about the opponent’s arguments. These patterns of interaction, in turn, are said to promote social support and safety, creativity, performance, and highly quality solutions to which participants are highly committed (Tjosvold, 1998).

Although some scholars, such as Allington (2002) and Shepard (2000), believe that high-stakes accountability further discourages the use of intellectual conflict among students, might a constructivist test preparation approach that involves constructive conflict enhance students learning processes? And while Johnson and Johnson (2009) argue that discussions spilling into chaos are particularly associated with lower-socioeconomic status and
lower-achieving students, might a constructive conflict approach assist these students, those who actually need test preparation the most?

In cooperative situations, distributed knowledge and different perspectives tend to be viewed as complementary and interdependent, which, in turn, increases accurate perspective taking, reduces competence threat, and focuses participants' attention on coordinating different points of view to enhance the cooperative effort, all of which tend to enhance learning and productivity (Butera, Huguet, Mugny, & Prez, 1994; Butera, Mugny, & Buchs, 2001; Gruber, 2006). Such an approach will be utilized during the classroom discussions in this study.
CHAPTER 3

METHODOLOGY

This study investigated the metacognitive processes (metacomprehension, problem solving, critical thinking) of students as they prepared for the CAHSEE exam. Specifically, it captured various “snapshots” of student thinking over a period of 3 months as they completed three, 25-question Princeton Review English language arts CAHSEE practice tests and reflected on their responses through classroom discussions and journaling. Group analyses allowed the researcher to determine what, if any, achievement level, a primary factor in the high-stakes test literature, played in this particular setting. In addition, six focus students participated in think alouds during testing and they participated in individual interviews at the conclusion of the study to help the researcher evaluate her approach as students explained what activities they found least and most helpful.

RATIONALE FOR METHOD

This study was a mixed methods action research study. While a quantitative approach captures scientific data of participants and may reduce the risk of researcher bias, it lacks the collaborative dialogue and participatory decision-making that a qualitative approach can bring (Denzin & Lincoln, 2005). Therefore, a study that includes both quantitative and qualitative data allows for maximum possibilities to not only report participant performance, but to investigate how to improve it, a fundamental aim of education.
Action research aims to "solve pertinent problems in a given context through democratic inquiry in which professional researchers collaborate with local stakeholders to seek and enact solutions to problems of major importance to stakeholders" (Greenwood & Levin, 2005, p. 54). According to this definition, the professional researcher could be described as the teacher researcher while the stakeholders might be seen to be the students, those needing to pass the CAHSEE. The approach for action research is both necessary and relevant since passing the CAHSEE is now necessary in order for students to graduate from high school.

A mixed methods approach is most effective because the actual CAHSEE ELA exam is quantitative in nature (with the exception of a qualitatively marked essay worth 20% of the total score). Therefore, students should be subjected to quantitative simulations like the actual testing situation with the opportunity to reflect on their numeric scores over time. Such an approach has been shown to improve student scores (Pressley et al., 1997). In addition to simulated practice tests, students should be encouraged to discuss and learn from their practice results (Kitsantas, 2002). A participatory research approach includes the added benefits of "shared ownership of research projects, community-based analysis of social problems, and an orientation to community action" (Kemmis & McTaggart, 2005, p. 561). Shared ownership, collective analysis, and actions, or solutions, that promote change, are all relevant aspects of the qualitative side of this mixed methods approach.

Finally, the key aspects of participatory action research fit extremely well into this study and serve as an effective model. The spiral series of thought is shown in Table 2. This repetitive research cycle encourages a series of actions initiated by a classroom group, or participatory action group. The planned change might be seen as improved practice test
Table 2. The Spiral Series of Thought

<table>
<thead>
<tr>
<th>Planning a change</th>
<th>Acting and observing the process and consequences of the change</th>
<th>Reflecting on these processes and consequences</th>
<th>Replanning</th>
<th>Acting and observing again</th>
<th>Reflecting again and so on</th>
</tr>
</thead>
</table>


scores while experiencing classroom discussions might be understood as observing the process of test-taking in-depth and the laborious thinking that it involves. The reflecting on these processes and the thinking required might be best captured through independent student journals while the replanning, acting and observing again, and the reflecting again invites opportunities to practice what was learned in the next testing cycle. In addition to providing an effective framework, action research is well suited to a constructivist approach. Such an approach may benefit low-achieving students the most because this approach involves social processes in which students learn and benefit from each other.

RESEARCH QUESTIONS

The basic research questions that guided this study are:

1a. How, or in what ways, does a series of practice tests influence students’ readiness for the actual CAHSEE exam?

1b. How do student groups perform on the average relative to each other (advanced vs. basic, basic vs. below basic, advanced vs. below basic?)

1c. How does practice test achievement between more advanced and less advanced focus students compare?

2a. How, or in what ways, do a series of classroom discussions influence students’ readiness for the actual CAHSEE?
2b. Specifically, what verbal articulation and efficacy patterns can be observed in the focus students?

3a. How, or in what ways, do a series of journal reflections influence students’ readiness for the actual CAHSEE?

3b. Specifically, what written articulation and efficacy patterns can be observed in the student groups (advanced, basic, below basic)?

4a. How, or in what ways, do a series of think alouds influence students’ readiness for the actual CAHSEE?

4b. Which aspects of the intervention do the focus students perceive to be most and least effective and why?

BACKGROUND/PILOT WORK

The present study resulted from a pilot study that the researcher conducted in a similar setting during the 2006-2007 school year. The pilot featured 92 ninth graders ($N = 92$) enrolled in one advanced, one basic, and one below basic English classes. Results suggested that while many students were able to talk about their metacognitive processes, including how they comprehended passages and how they solved specific test questions, most were not able to detect patterns in their approaches until they wrote about specific test items over a period of several months. It was only then that they developed critical thinking skills, such as identifying patterns in their prior thinking. During these discussions, themes began to emerge. Students would identify patterns such as, “I read something into the question that wasn’t there,” “I didn’t see the key words in the stem of the question,” “I didn’t narrow down the selections to two choices,” and eventually, after many sessions, “Now I can see through the question.”

While the intervention appeared to elevate students’ awareness and insight about what they needed to do to be successful on the CAHSEE, it was impossible to tell if this were true
for specific student groups since only one large, 72-item multiple-choice test was administered at the beginning of the school year, a test that was dissected through monthly classroom discussions for a period of roughly 6 months. No matter how many discussions or journal entries students completed, the pilot was not equipped with any type of posttest to measure progress. The only conclusive data were field notes taken by the teacher researcher after each classroom discussion, journal entries in which some students identified improvement patterns, and informal interview testimonials in which a few students claimed that the intervention really helped them to be more successful on standardized tests. As one student put it, “to see through the test.”

Although no direct conclusions could be established regarding the intervention, students who experienced the pilot appeared to perform better on the actual CAHSEE. Of the approximately 145 students who took the CAHSEE for the first time in March 2008, 33 students, or 22.7%, did not pass the English Language Arts portion of the CAHSEE. Of the 95 students who participated in the pilot, 7, or 7.3%, failed the ELA portion compared with the remaining group of 50 students who experienced a failure rate of 52% with 26 students failing the ELA portion. Among those 7 who failed and received the intervention, 4 were ELLs, 1 was a SPED student, and 2 were regular education students. Of the 7, 6 were male and 1, a regular student, was female.

The results from that study left many lingering questions, including, did the basic or below basic English classes benefit more from the intervention than the advanced class, a question that previous research has suggested is typically not the case (Kulik et al., 1984; Vernon, 1954). While Pressley et al. (1997) theorized that test-taking skill is a byproduct of information processing capabilities, could it be possible to provide an intervention that
improves students', particularly low achieving students', information processing capabilities in between admissions of tests? This pilot study influenced this researcher and motivated her to refine her methodology so that a similar, more detailed and thoughtful study could be replicated in the form of a dissertation.

THE SITE AND THE PARTICIPANTS

The site was a high school located in an urban district in southern California with a 2009 academic index performance (API) score of 713. The high school is considered to be a "small school." Beginning in the 1990s, major reforms of comprehensive high schools have been enacted in order to create more personalization and greater success among low-income students (Darling-Hammond, 2006). With the assistance of the federal government through the Small Schools Act, several schools throughout the United States have taken part in this phenomenon. In addition, The Bill and Melinda Gates Foundation has provided funds for these types of schools. In 2004, this large, comprehensive high school was broken into four separate schools, now containing approximately 450 students in each school. This site was chosen primarily because the researcher is a teacher of English at this location. However and in addition, small schools are ripe for research because they are experimental in nature. According to 2009 STAR testing data, 67% of the population tested is socioeconomically disadvantaged, 30% are English Language Learners, and 11% have disabilities (CDE, 2009). In addition, students in this small school are racially and economically diverse, as Table 3 indicates.

For the 2008-2009 school year, the district superintendent imposed a size limit on freshman English classes: 20 students per class. Thus, the 60 students eligible for this study
Table 3. Ethnic/Racial Academic Performance Index Subgroup Information in the Small School Where This Study Took Place

<table>
<thead>
<tr>
<th>Student Group</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>African American</td>
<td>71</td>
<td>19</td>
</tr>
<tr>
<td>American Indian</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Asian</td>
<td>45</td>
<td>12</td>
</tr>
<tr>
<td>Filipino</td>
<td>29</td>
<td>8</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>136</td>
<td>39</td>
</tr>
<tr>
<td>Pacific Islander</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>White</td>
<td>68</td>
<td>19</td>
</tr>
<tr>
<td>Two or More Races</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>359</td>
<td>100</td>
</tr>
</tbody>
</table>


represented three classes, advanced English (n = 20), basic English (n = 20), and below basic English (n = 20). In the advanced class, two of the students were English language learners but had no special needs. In the basic class, none of the students were English language learners but two had special needs. Finally, in the below basic class, two of the students were English language learners and one had special needs. Combined, the three classes comprised about 40% of the freshman class (N = 150). In past years, the average English class size has been about 30. Therefore, the current advanced English class was unable to absorb all of the advanced students so some regular English classes now contained about 10 advanced students per class of 20. This made these classes less homogenous as the range of students was greatest in these types of classes.

The student groups, or classes, for this study were randomly assigned to the teacher researcher. One group was advanced and tended to perform well on standardized tests (n =
This group stayed focused and on-task through virtually all assignments. Nine students from this group were included in this study. Students enrolled in advanced English are considered to be part of the school’s Gifted and Talented Education (GATE) program. High achieving students, regardless of their income level, tend to exercise characteristics such as self-reliance and resilience (Walker, 2007). These types of students tend to be able to predict, rather accurately, whether or not they know specific testing material whereas underachievers, often students who have low GPAs, do not have the same clarity about what they know or do not know (Hong et al., 2006; Kim & Goetz, 1993; McClain, 1983).

The next group of students in this study was considered basic (n = 20). However, due to the new configuration of students per class, this group contained a significant amount of high achievers. Seventeen students from this group were included in this study. This class could be considered, collectively, to be moderate achievers, in between the high and low achieving classes. Of all the groups, they had the greatest range of skills.

The final group of students in this study was considered low-achieving (n = 20) because they tended to test below basic on standardized tests. Eleven students from this group were included in this study. These types of students tend to struggle with basic literacy and have had less success in school. Such students tend to exert less effort and persist for shorter amounts of time on tasks (Bandura & Cervone, 1983, 1986, cited in Crocker, 2005; Zimmerman, 1995). They also tend to set lower goals and are less motivated (Locke & Lathan, 1990, 1994, cited in Crocker, 2005; Zimmerman & Bandura, 1994, cited in Crocker, 2005).
DATA COLLECTION

This section describes the instruments used and procedures followed in this study.

Instruments

The instruments that were used in this study are briefly described below. A more thorough explanation of how each instrument was used can be found in the procedures section that follows.

INFORMED CONSENT FORMS

The parental letter found in Appendix A details the nature of the study, risks, benefits, and clearly states that participants can opt out of the study at any time. A similar but less detailed assent form was constructed for the students involved in this study (see Appendix B). In addition, in order for students to be considered as focus students, additional consent was needed including permission to perform recorded think alouds and to participate in recorded interviews.

PRINCETON REVIEW TESTS

The Princeton Review provides tests that teachers may administer to prepare students for the CAHSEE. The Princeton Review tests used for this study were a reconfiguration of two English language arts tests, test 1 and test 2, consisting of approximately 16 reading selections and 72 multiple-choice questions (Princeton Review, 2005). For this study, the Princeton Review revised practice CAHSEE tests, forms A, B, and C, were used. For a detailed explanation of how these tests were constructed, please see the explanation in the data procedure section. In addition, see Appendix C.
CLASSROOM DISCUSSIONS

The in-class discussions, which occurred two times in between each monthly test administration, consisted of student volunteers explaining their metacognitive processes (metacomprehension, problem solving, critical thinking) as they approached particular test questions. They discussed three, class-selected questions to explore in depth during a 30-minute period. These discussions were audiotaped.

JOURNAL REFLECTIONS

At the conclusion of each 30-minute classroom discussion, students wrote a one-half page reflection on one of the three test questions explored. This test question was selected by the teacher researcher in order to ensure more accurate analysis because the data could be analyzed according to the same prompt. The reflection required students to extrapolate on some of the test-taking strategies explored during the classroom discussion (see Appendix D).

FOCUS STUDENTS PARTICIPATING IN THINK ALOUDS

During the test-taking process, two students from each class, students who elected to participate as focus students, explained their thinking process out loud while they talked into an audiotape recorder. They were allowed to test in a private setting (library) under the direction of a librarian who monitored their participation and encouraged them to talk into the tape recorder while testing.
**End of Study Interviews**

At the conclusion of the study, each of the six students, the same students who were selected as focus students, participated in one, 20-minute interview about what aspects of the study were least and most helpful (see Appendix E). In addition, they were asked about the think aloud process and they were able to review their previous think aloud transcripts to gain insights as to how their thinking may have changed throughout the test-taking process.

**Procedures**

The procedures that were used in this study are described below.

**Informed Consent Forms**

Informed consent parental permissions were sent home to each parent/guardian (see Appendix A) and returned via student. Parents who did not return the letters after 1 week were called. If verbal permission was granted for the study via phone, students were given another form, asked to deliver it to his her parent(s)/guardians(s) and then to return it to the teacher researcher. Any parents who did not grant permission for their child to participate in the study and/or did not return the permissions form could not have any of their child’s data included in the study. In addition, students who did not wish to participate in the study could not have any of their information included in the data set. In addition, only parents and children who wished to participate in the additional role of producing think alouds were considered to be focus students. Any child who was selected as a focus student and wished to decline during any time during the study could do so. See Appendices A and B for a full account of permissions.
PRINCETON REVIEW TESTS

The Princeton Review provides tests that teachers may administer to help students prepare for the CAHSEE. Each test, test 1 and test 2, consists of approximately 16 reading selections and 72 multiple-choice questions. Most reading selections comprise of between two and six reading comprehension questions relative to each passage. Passage selections vary. The classifications of reading passages are literary texts, poetry, informational science texts, informational social science texts, and functional documents. Sections on grammar are also considered functional documents even though these sections are not accompanied by a text.

A weakness of the pilot study was that it did not include a follow-up or posttest after the intervention. While it might be possible to now give test 1 as a pretest and test 2 as a posttest, this would not be a reliable method because the two tests are not parallel. While they do contain many of the same types of selections, they are not equally weighted from one test to the next. For example, each test contained a long section. However, one test contained 9 comprehension questions from a literary text while another contained 12 comprehension questions from an informational science text. In addition, when administering the 72 questions test prior, it took students approximately 2½ hours to complete it. Many considered the task laborious and lost interest. Clearly, a more effective approach would require shorter tests administered between the intervention of classroom discussions and journal reflections to see if such interventions were making a difference over time.

In order to create three smaller tests and to make them as precise as possible, all reading selections were categorized according to Princeton’s labeling system of texts. Reading selections containing poetry were omitted since there were only two selections, one
on each test. In addition, passages that contained extensive amounts of questions, seven or more, were also eliminated since one test contained a long series of questions from a literary text while another contained a long series of questions from a science informational text. Finally, questions about grammar were omitted because they were not part of larger readings.

Each group was given a different test during each cycle of the study. For example, the advanced group was given Test B, the basic group was given Test A, and the below basic group was given Test C at the onset of the study. This allowed for a counterbalance test design to ensure greater reliability (Isaac & Michael, 1995; Vernon, 1954). Four weeks later, the advanced group was given Test A, the basic group was given Test C, and the below basic group was given Test B. Four weeks after this time, the advanced group was given Test C, the basic group was given Test B, and the below basic group was given Test A. Students had approximately 60 minutes to complete each 25-question test. Students who needed more time were able to complete the test the next day. Students who were absent also completed the test upon their return. See Appendix C for a complete outline of the newly revamped, 25-question tests. These were administered approximately every 4 weeks for 3 months during a semester with discussions and journal reflections occurring for two cycles in between test administrations (see Figure 1 for a visual guide).

CLASSROOM DISCUSSIONS

Two weeks after the first testing onset, each group was given a print-out of how the class did on the first three reading selections of their test. The focus was on classroom progress, not on individual progress. All student names on the print-outs were replaced with the last four digits of each student’s identification number to ensure confidentiality and to
Figure 1. The cycle model: Testing, discussing, reflecting, discussing, reflecting, testing.

keep the focus on group progress. Students added up the number of correct answers per test item (see Appendix F).

After voting by secret ballot about which test question each group would like to review per reading selection (three test questions total), students discussed each test question in-depth during 30-minute classroom discussions that were transcribed by a professional transcriber. The discussion transcripts were also checked against the original recordings and proofread so that any errors could be detected before analysis began. Following each discussion, field notes were taken on general observations and impressions of the verbal and non-verbal behaviors of the participants during the discussions. Data collection and analysis occurred concurrently in order to continually refocus and refine the study.
During pilot work, students typically selected the test question that the class did the poorest on for each reading selection. For example, in Appendix F, which details the reading selection, “A Visit to the Doctor’s Office,” the students scored the worst on question 3, with only 20.8% of the class getting this item correct. Typically, a class would choose this test question for an in-depth review. This decision to review the poorest performing question was consistent among all three classes. For this study, the three test questions that the students voted to explore became the focus of the classroom discussions for the next 30 minutes. In addition, think aloud data on these questions from the six focus students were studied closely.

Initially, if students did not volunteer to discuss a test question, the teacher researcher modeled her own metacognitive processes on a student-selected test item (see Appendix G). For example, the discussion might have gone like this:

Only 5 out of 24 of us got question #3 correct from the reading, “A Visit to the Doctor’s Office.” Now question #3 reads:
This selection is BEST described as--
A. fiction
B. biography
C. article
D. drama

Now when I look at this question, I first have to think about what each one is. What is fiction? What is a biography? What is an article? How is an article different than fiction or a biography? What is a drama? How is a drama different than fiction? Maybe I should start by figuring out which ones are non-fiction and which ones are fiction [teaching classification of the answers]. Okay, A is fiction because it says it’s fiction. I know B is non-fiction because a biography is a true story about someone. I know C is non-fiction. Articles in newspapers are typically non-fiction. They are true. I know D is fiction because drama is a play and that is a type of fiction. So now I know that A and D are fiction and B and C are non-fiction. I have now narrowed down, or classified, the answers into two groups.

The question asks what the selection can be best described as. Best is capitalized so the test maker is looking for the best answer. That means other answers may also be true but I need to find the best one, or the one that’s true more than any other. By looking back at the beginning of the story, I can see that the test maker tells me that, “This is a story about a young woman and the
uneasiness she feels when visiting the doctor’s office.” Based on that description, I can deduce, or figure out, that this reading is a story and a story is typically fiction so the answer must be either A “fiction” or D “drama.” Now I know that a drama is written like a play with the names of characters capitalized followed by the words that they say and this text does not look like that. I decide the answer is A, “fiction.” Now can anyone explain why he/she thought it was C because I can see from the print-out that 10 of you thought it was C, “non-fiction?” I can also see that 8 of you thought it was B, “biography.” Can anyone who selected those answers explain why?

If students do not volunteer, I will be a hypothetical student stating something like, “I thought it was C, ‘non-fiction,’ because I thought the story was true.” Or “I thought it was B, ‘biography,’ because I thought it was a true story about someone.” Then I’d explain that no matter what I thought, the test maker has actually told me that it is a story. Therefore, the best answer is A, fiction. I’d also ask volunteers to explain how we know it’s a story. “How is it written? Is there dialogue? Is there description?” “How is the dialogue written?” “Do characters talk to each other?” “How is that dialogue different than dialogue you might find in an article or a biography or a drama?” “Why is A the best answer?”

Students repeated the classroom discussion process approximately 1 week after the initial discussion on the last three reading selections of the test, 2 weeks before the next test was given. Classroom discussions repeated twice in-between each test administration, a total of three times, for a total of six discussions per class.

**JOURNAL REFLECTIONS**

At the conclusion of each classroom discussion, students wrote a one-half page reflection on one of the three test questions. The teacher researcher selected the test question so that the data could be analyzed from the same prompt. Students followed the journal prompt detailed in Appendix D. They first wrote down the class test question. They then
explained how they approached the test question initially, and they then explained how they might approach the test question differently based on our classroom discussion. For example, the conversation went something like this:

After our discussion on question #3 from “A Visit to the Doctor’s Office,” first rewrite the question and then explain why you missed the test question. [Students who got the question correct would explain how they think someone might have missed it.] The second part of the question is, if you were to see this question on a test again, how you might respond to it differently. [Students who got the question correct would explain how they think someone might respond to it differently if they missed it.]

Think aloud data from focus students was carefully compared to classroom discussion information (see Figure 1 for a visual guide of the data collection procedure).

Because writing about thinking processes may prove to be more challenging for students than merely talking about them, the teacher researcher needed to model this approach. The modeling went something like this.

Here is the test question, #3, from “A Visit to the Doctor’s Office.”

This selection is BEST described as--
A. fiction
B. biography
C. article
D. drama

I first missed the test question because I was thinking about the story and I thought that since it was about a girl named Janelle who was afraid to go to the doctor’s office, it was a true story. I thought it was about her life. I didn’t realize that the test maker told me in the beginning that it was a story. Also, after our discussion, I now see that if it really was a biography, it would be more like the writer telling me about this girl. It wouldn’t have her being the main character with the doctor talking to her like that. Instead, the writer of the biography would be explaining things about Janelle but not in her own words like this.

If I were to see this test question again, I would do what we did in class. I would try to narrow it down to two answers and then work from there. I would try to classify the answers too so I could figure out the right one. I would also pay more attention to the clues the test maker gives like capitalizing best so I know I have to pick the answer that is the most correct even though other answers might come close.
This journal reflection process repeated each time after a classroom discussion took place. During the reading of the journals, the researcher made a list of critical thinking behaviors that continually appeared in the journal writings of students (i.e., paraphrasing the test question, seeking confirmation of an anticipated answer in the text, and eliminating wrong answers with textual rationale). These themes were briefly discussed with students as a whole class each time they received their journals back. All journals were examined within a few days of students' responding. Written feedback was given to the students regarding how clearly they answered the prescribed questions.

**FOCUS STUDENTS PARTICIPATING IN THINK ALOUDS**

In order to get a closer look at what, specifically, students were thinking during and after the test-taking cycle, six students were selected to serve as focus students. These students were selected purposively, reflective of the two largest demographic groups of the overall student population, Hispanic and African American, involved in this study. Because Hispanic and African American students are most at risk for failing the exam according to current CAHSEE ELA data, it is necessary to glean information from these populations. In addition, test questions that whole class groups elected to review could be compared with think aloud data for those same test questions.

Of the 37 students who participated in the study, 12 are Hispanic (34.42%), 12 are African American (34.42%), 7 are Asian (18.91%), and 6 are White (16.21%). Of the 37 students in the study, 17 are female (45.94%) and 20 are male (54.05%). Focus group data were compared to quantitative mean data generated for the entire sample. Were trends observed in the larger group echoed in the focus group students? (See Table 4 for additional
Table 4. Demographics of Focus Students

<table>
<thead>
<tr>
<th>Type of Test-Taker</th>
<th>Name</th>
<th>Gender</th>
<th>Ethnicity</th>
<th>8th Grade English CST Score</th>
<th>8th Grade English CST Level</th>
<th>ELL?</th>
<th>SPED?</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Rashawndra</td>
<td>Female</td>
<td>Afr. Amer.</td>
<td>455</td>
<td>Advanced</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>High</td>
<td>Nathaniel</td>
<td>Male</td>
<td>Afr. Amer.</td>
<td>437</td>
<td>Advanced</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>High</td>
<td>Bill</td>
<td>Male</td>
<td>Afr. Amer.</td>
<td>393</td>
<td>Proficient</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Low</td>
<td>Reynaldo</td>
<td>Male</td>
<td>Hispanic</td>
<td>328</td>
<td>Basic</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Low</td>
<td>Alejandro</td>
<td>Male</td>
<td>Hispanic</td>
<td>280</td>
<td>Below Basic</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Low</td>
<td>Jenaya</td>
<td>Female</td>
<td>Afr. Amer.</td>
<td>247</td>
<td>Far Below Basic</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*Names of all students have been changed to protect their privacy.

Specific student selections were based on parent and student willingness to give consent as noted on the student permission form (see Appendix A). In addition, student willingness to participate was also granted as noted by marking the focus student box on the student permission forms. Once these two requirements were met, individual focus students were selected based on California State Standards test results for English Language Arts from 8th grade. Those who scored “advanced” or “proficient” were considered high test-takers. Those who scored “basic,” “below basic,” or “far below basic” were considered low test-takers. This demarcation is in accordance with the aims of NCLB law that calls for 100% of students testing at proficient or above levels by 2014.

END OF STUDY INTERVIEWS

At the end of the study, the six focus students, the same students who took part in the think alouds, were selected to participate in interviews about what aspects of the intervention details.) In addition, the researcher decided to include three high and three low test-takers, since low test-takers are an especially vulnerable group for not passing the CAHSEE and/or dropping out. This way, the behaviors of the two groups could be compared.
were the least and most helpful (see Appendix E). This data was compared with the quantitative and qualitative data conclusions previously drawn and made an interesting comparison as to whether or not student opinions confirm or deny focus student levels of progress. In addition, it gave individuals opportunities to expound on potential causes for test performances.

These interviews (approximately 20 minutes each) occurred during the regular classroom hour and occurred while a substitute teacher was present to ensure the manageability of the remaining students. These interviews were digitally recorded in a private setting (library), one student at a time. They were then transcribed verbatim by a professional transcriber. The interview transcripts were checked against the original recordings and proofread so that any errors could be detected before analysis began. Following each interview, field notes were taken on general observations and impressions of the non-verbal behaviors of the participants during the discussions.

**DATA ANALYSIS**

Data analysis included a review of all types of data collected (Princeton Review test scores, classroom discussions, journal entries) with a particular focus on the six focus students (think alouds during testing, end of study interviews).

**Quantitative Analysis**

The primary quantitative means of analysis was through examining The Princeton Review test scores over three points in time to see if growth and/or change could be observed in the group classes (advanced vs. basic, basic vs. below basic, advanced vs. below basic). All test question data was entered in SPSS 16.0 for the Macintosh, a statistical software
package. First, a general linear model ANOVA was conducted to glean an overall impression of the student group performances. Then, repeated measures were conducted to investigate differences relative to group performances on all of the three tests (March, April, May). Finally, pairwise comparisons were conducted to see if there were differences among the tests since they were given to the classes in different orders. Collectively, these tests explain changes in the average mean scores over time and they suggest why some groups may have outperformed others. Such tests allow the researcher to see if the intervention was more influential for certain classes than others. For example, did the basic class have a higher mean improvement score than the advanced class? Besides this cursory quantitative approach that only gives a general picture of student performance, the main analyses in this study were qualitative in nature and focused on classroom discussions, journal responses, and focus student think alouds.

**Qualitative Analysis**

For the qualitative aspects of the analysis, all transcripts (focus student think alouds, classroom discussions, focus student interviews) were reviewed by coding the data. Through a grounded theory approach (Glaser & Straus, 1967; Strauss & Corbin, 1990), codes were established via a constant-comparative approach, first through analyzing the focus student think aloud data individually as it unfolded each time they tested. Then, classroom discussion data transcripts were analyzed. Transcript analysis involved reading transcriptions from the general classroom discussions each time they took place and studying the content, categories, and context. They were compared to the themes emerging in the focus student data. Throughout this repetitive process, the researcher consulted her individual field note memos.
consisting of impressions and evolving conclusions as the study unfolded. The field note memos, combined with the analysis procedure, created a looping process in which describing, classifying, and interpreting (Creswell, 1998) could become apparent.

Emerging themes were then extracted beginning with impressions from think alouds, and then classroom discussions, and finally journal entries. Journal entries were primarily used to either validate or negate findings established from the classroom discussion transcripts. The constant comparative method (Glaser & Strauss, 1967) involved coding an occurrence for a category and then comparing it with previous occurrences. This cyclical, looping phase fit in well with the repetitive nature of cycles within this study.

Like the focus student think alouds and the in-class discussions, the interviews were analyzed through a grounded theory approach (Glaser & Strauss, 1967; Strauss & Corbin, 1990). Codes were established via a constant-comparative approach, first through analyzing the individual interview data and then comparing emerging themes among the focus students. This interview data was analyzed within 2 weeks of the given June interviews. Finally, a chronological analysis was conducted throughout the 3-month data collection period in which certain themes examined at various points in time continually reappeared over the duration of the data collection process.

In-depth comparisons between and among the six focus students were conducted using a coding system (Glaser & Strauss, 1967). First open coding was employed. All of the documents were reviewed for salient categories of information. Once categories were established, data was then reviewed until a category was saturated. Next, a central category, axial coding, was identified. Interrelationships between categories were examined, including a review of conditions that may have influenced the central category or categories. Because
not all categories are equally relevant, the depth of inquiry into each one varied (Glaser & Strauss, 1967). The "story" that connects the categories could then be established (selective coding) and theory could be tied in.

Glaser and Strauss (1967) recommend that substantive theory come before formal theory. This research allows the researcher to code data, make connections, and form ideas without being constrained by previous literature. Substantive theory, theory that is developed for a specific set of inquiry and that is closer to real-world application, tends to reveal aspects of formal theory which is more general, abstract, and conceptual. For this research study, substantive theory emerged through the unveiling of students' metacognitive processes under the premise that individual aspects of it (i.e., think alouds, discussions, and reflections) serve to enlighten student knowledge and consequently, performance on standardized tests. In addition, formal theory about motivation and self-regulation, had important implications in this study.
CHAPTER 4

RESULTS

This chapter will include both the quantitative and qualitative results from this study. First, an overview of quantitative data will be presented including a one-way analysis of variance (ANOVA) on the total scores of earned by each group, advanced, basic, and below basic. This will allow for an overall picture. Next, the analysis will be refined through a repeated measures ANOVA by individual classes over time to determine the differences, if any, between and among classes at all three testing points, March, April, and May. Finally, pairwise comparisons will be presented to determine if mean differences existed between the three practice tests given. Collectively, these results will reveal how the students performed quantitatively and suggest why some groups may have performed better than others. Finally, a summary of the key findings will be presented.

After the quantitative data is presented and serves as a backdrop, qualitative data will be presented by themes. These themes will explore students’ levels of articulation and efficacy and will be categorized in the following six ways: (a) reading the test question and answers, (b) interpreting the test question and answers, (c) eliminating answers, (d) revisiting the text, (e) integrating the above strategies concurrently, and finally (f) expressing feelings about test-taking abilities.
INTRODUCTION TO THE QUANTITATIVE TEST RESULTS

The following section details results of the data collected to measure student growth on practice CAHSEE standardized tests over time for advanced, basic, and below basic classes. The researcher’s driving question was: How, or in what ways, does a series of practice tests influence students’ readiness for the actual CAHSEE exam? Assisted by the Princeton Review test preparation materials, the researcher developed three 25-question tests with an equal number of science, social science, literature, and functional document readings (see Appendix C).

Overall Mean Test Results for Advanced, Basic, and Below Basic Classes

The overall results indicated that there were differences between group performance levels as evidenced by the total means scores for all three tests (see Table 5). Each test consisted of 25 items with a grand total of 75 items. These results clearly indicate that the advanced students scored the highest (57.64), the basic students scored the next highest (47.59), while the below basic students scored the lowest (38.11). When examining the ranges of the groups, it becomes apparent that the basic group had the largest range, with a low score of 26 and a high score of 68, a range of 42 points. This group also had the greatest standard deviation (11.375) meaning that 68% of the students in this class had scores in the range of 36 to 60 correct while 95% had scores in the range of 25 to 70 correct. This suggests that there was greater variability for this class and this makes sense since, as outlined in the methodology section, this group was a hybrid of low and high performing students.
Table 5. Mean, Standard Deviation, and Range for All Three PR Practice Tests (Combined)

<table>
<thead>
<tr>
<th>Classes</th>
<th>$N$</th>
<th>Mean</th>
<th>$SD$</th>
<th>SE</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Minimum Range</th>
<th>Maximum Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced</td>
<td>11</td>
<td>57.64</td>
<td>7.173</td>
<td>2.163</td>
<td>52.82</td>
<td>62.46</td>
<td>46</td>
<td>69</td>
</tr>
<tr>
<td>Basic</td>
<td>17</td>
<td>47.59</td>
<td>11.375</td>
<td>2.759</td>
<td>41.74</td>
<td>53.44</td>
<td>26</td>
<td>68</td>
</tr>
<tr>
<td>Below Basic</td>
<td>9</td>
<td>38.11</td>
<td>9.597</td>
<td>3.199</td>
<td>30.73</td>
<td>45.49</td>
<td>27</td>
<td>51</td>
</tr>
<tr>
<td>Total</td>
<td>37</td>
<td>48.27</td>
<td>12.045</td>
<td>1.980</td>
<td>44.25</td>
<td>52.29</td>
<td>26</td>
<td>69</td>
</tr>
</tbody>
</table>

Unlike the basic group, both the advanced and the below basic groups tended to score more similarly within their groups although their performances were vastly different when compared to each other. For example, the advanced class had the smallest range, with a low score of 46 and a high score of 69, a range of 23 points, nearly half that of the basic group. This group also had the smallest standard deviation (7.173) meaning that 68% of the students scored in the range of 50 to 65 correct and 95% had scores in the range of 43 to 72 correct. The below basic group had a similar range, a range of 24 points, but they scored considerably lower than the advanced group. For example, the lowest score was 27 and the highest was 51. The below basic group had a standard deviation of 9.597, meaning that 68% scored in the range of 29 to 48, while 95% scored in the range of 20 to 58. Although the advanced and below basic groups scored differently with respect to the range within each group, they were more homogeneous than the basic group that had a very wide range.

After conducting a one-way ANOVA of overall scores by classes, the overall post hoc Tukey $F$ test, a liberal test, indicated a significant difference between the advanced group and the basic class ($p = .033$) and an even greater difference between the advanced group and the below basic class ($p = .000$). The overall post hoc Scheffé $F$ test, a more stringent test, also
indicated similar results. There was a significant difference between the advanced class and the basic class \( (p = .043) \) and between the advanced group and the below basic class \( (p = .000) \). These results suggest that the advanced group was significantly different than the other groups studied. However, it was not clear how. Therefore, other tests, specifically tests exploring the individual class means of each test over time were needed.

**Individual Class Mean Test Results on Each Test Over the Course of the Study**

Although each class had different patterns with respect to their performances over time that will be discussed shortly, it is important to note some common patterns among all three groups. Although mean scores varied from group to group, all groups had the lowest mean score on Test B, regardless of when they took it. This is particularly interesting to note since the three classes took the tests in different orders. The advanced class that took Test B in March had a mean score of 18.081, the basic class that took Test B in May had a mean score of 15.235, while the below basic class that took Test B in April had a mean score of 12.222 (see Table 6). Coincidentally, the standard deviation scores for all groups were also greatest on Test B with the advanced group having a standard deviation score of 3.885, the basic group having a standard deviation score of 5.426, and the below basic group having a standard deviation score of 4.790. Since the highest range of scores for all groups was on Test B, this suggests that Test B was an inferior test to the others because it produced greater variability.

In addition to all groups scoring the lowest on Test B, all groups scored the highest on Test A regardless of when they took it. The advanced group who took Test A in April had a mean score of 20.727, the basic group who took the test in March had a mean score of
Table 6. Mean and Standard Deviation Scores of Individual Classes Over Time

<table>
<thead>
<tr>
<th>Test</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
<th>Test</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
<th>Test</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>18.818</td>
<td>1.991</td>
<td>11</td>
<td>B</td>
<td>15.235</td>
<td>5.426</td>
<td>17</td>
<td>A</td>
<td>13.333</td>
<td>4.416</td>
<td>9</td>
</tr>
</tbody>
</table>

16.353, while the below basic group who took the test in May had a mean score of 13.333. However, unlike Test B, in which all groups showed the highest standard deviation, or variance, the results from Test A don’t yield the smallest standard deviation with the exception of the basic group that took it first in March (3.297). The other groups, advanced (3.165) and below basic (4.416), experienced their midpoint standard deviation through Test A, another indication that they were more homogeneous than the basic group. Therefore, it occurred to the researcher that other factors might have affected student performance besides the degree of difficulty that the tests held for the students. The basic group showed a consistent pattern of having a lower standard deviation than their other tests (3.297), but there may have been other reasons for their variance than the test itself; otherwise, their standard deviation scores should have decreased over time. Prior to that discussion, it is first imperative to report out the individual mean and standard deviation scores for each class over time.

**ADVANCED MEAN AND STANDARD DEVIATION SCORES OVER TIME**

In order to get a clearer picture of the performance of each group individually over time, the data were analyzed by each class. As Table 7 indicates, the advanced class, which completed Test B in March, Test A in April, and Test C in May, experienced an initial
Table 7. Mean and Standard Deviation Scores of the Advanced Class Over Time

<table>
<thead>
<tr>
<th>Test B (March)</th>
<th>Test A (April)</th>
<th>Test C (May)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>18.091</td>
<td>3.885</td>
<td>11</td>
</tr>
<tr>
<td>20.727</td>
<td>3.165</td>
<td>11</td>
</tr>
<tr>
<td>18.818</td>
<td>1.991</td>
<td>11</td>
</tr>
</tbody>
</table>

increase and then a decrease in their mean scores on the tests. On their first attempt, the advanced test-takers’ mean score was 18.091, then increased to a mean of 20.727, and then decreased to a mean of 18.818, just slightly higher than their first attempt. Interestingly, even though their scores decreased from their second to third attempt, their standard deviation scores decreased, indicating that the range, or variance of the student test scores decreased as the study went on, declining from 3.885 on the first test down to 1.991 by the last.

Results for the ANOVA for the advanced group indicated a significant testing effect (Wilks’ Λ; p = .050). Important to understand, however, is that this effect was from the initial test in March to the mid-study test in April. The change in scores from 18.091 to 20.727, an average gain of 2.636, was significant. However, this may have had to do with the fact that Test B was an inferior test while Test A appeared to be superior (all groups performed best on this test). The advanced group was the only group that experienced a series of taking Test B in March and then Test A in April. No other group had that configuration although the below basic group did take Test B in April and Test A in May and experienced an average gain of 1.111. However, this was not enough of a gain to be considered significant. The advanced groups’ significant testing effect could have been influenced by the order in which they tested. For example, they may have gained the most from March to April and gained less
from April to May. As the focus student data details, many students seemed to experienced a score decline from April to May (see Table 8).

Table 8. Focus Student Results on Overall Practice CAHSEE Tests

<table>
<thead>
<tr>
<th>Name of Focus Student and Prior ELA/CST Classification from 8th Grade</th>
<th>Name of English Class Student Was In</th>
<th>Practice CAHSEE Tests</th>
<th>Total Number Correct Out of 75 on All Practice CAHSEE Tests</th>
<th>Total Percentage Correct Out of 75 on All Practice CAHSEE Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nathaniel⁴ &quot;Advanced&quot;</td>
<td>Adv.</td>
<td>March: 23(B)</td>
<td>April: 25(A)</td>
<td>May: 20(C)</td>
</tr>
<tr>
<td>Bill⁴ &quot;Proficient&quot;</td>
<td>Adv.</td>
<td>March: 22(B)</td>
<td>April: 20(A)</td>
<td>May: 19(C)</td>
</tr>
<tr>
<td>Rashawndra &quot;Advanced&quot;</td>
<td>Basic</td>
<td>March: 19(A)</td>
<td>April: 19(C)</td>
<td>May: 23(B)</td>
</tr>
<tr>
<td>Alejandro &quot;Below Basic&quot;</td>
<td>Basic</td>
<td>March: 10(A)</td>
<td>April: 9(C)</td>
<td>May: 11(B)</td>
</tr>
<tr>
<td>Reynaldo⁴ &quot;Basic&quot;</td>
<td>Below Basic</td>
<td>March: 12(C)</td>
<td>April: 14(B)</td>
<td>May: 9(A)</td>
</tr>
<tr>
<td>Jenaya⁴ &quot;Far Below Basic&quot;</td>
<td>Below Basic</td>
<td>March: 10(C)</td>
<td>April: 11(B)</td>
<td>May: 8(A)</td>
</tr>
</tbody>
</table>

⁴Experienced a decline from April to May.

The overall trend of the focus students was a decline in scores from April to May and this can be seen in two of the three classes, the advanced group and the below basic group, the groups that were more homogeneous. Interestingly, the focus students in the basic group actually increased their scores from April to May. During this time, they went from taking Test C to taking Test B, the most difficult test. However, overall data on the basic class does not display this trend.
BASIC MEAN AND STANDARD DEVIATION SCORES OVER TIME

As Table 9 indicates, the basic class, who completed Test A in March, Test C in April, and Test B in May, experienced a consistent decrease in their mean scores on each test as the study went on. On their first attempt, the average test-takers maintained a mean of 16.353, then decreased to a mean of 16.000, and then ended with a mean of 15.235, lower than both their first and second attempts. Results for the ANOVA did not indicate a significant testing effect (Wilks’ $\Lambda; p = .470$).

Table 9. Mean and Standard Deviation Scores of the Basic Class Over Time

<table>
<thead>
<tr>
<th>Test</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test A (March)</td>
<td>16.353</td>
<td>3.297</td>
<td>17</td>
</tr>
<tr>
<td>Test C (April)</td>
<td>16.000</td>
<td>4.077</td>
<td>17</td>
</tr>
<tr>
<td>Test B (May)</td>
<td>15.235</td>
<td>5.426</td>
<td>17</td>
</tr>
</tbody>
</table>

Interestingly, even though the basic students’ scores decreased throughout the study, their standard deviation scores increased simultaneously. On their initial attempt in March, their standard deviation score was 3.297. It then increased to 4.077 on their April test, and finally, to 5.426 on their final test. Clearly, the students established higher standard deviation scores as the study went on, suggesting that the greater range of scores may have been due to the fact that the basic class was a hybrid class consisting of many high and many low test-takers. This could also indicate that many students lost interest and stopped trying. These ideas are explored further in the qualitative section.
BELOW BASIC MEAN AND STANDARD DEVIATION SCORES OVER TIME

As Table 10 indicates, the below basic class, which completed Test C in March, Test B in April, and Test A in May, experienced an initial decrease followed by an increase in their mean scores on each test as the study went on. On their first attempt, the below average test-takers’ mean score was 12.556, then decreased to a mean of 12.222, and then increased to a mean of 13.333 by the end of the study, just slightly higher than their initial mean score. Interestingly, their standard deviation initially was lowest at 3.087 on the March test then increased to 4.790 on the April test and then decreased again to 4.416 by the May test.

Table 10. Mean and Standard Deviation Scores of the Below Basic Class Over Time

<table>
<thead>
<tr>
<th>Test</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test C (March)</td>
<td>12.556</td>
<td>3.087</td>
<td>9</td>
</tr>
<tr>
<td>Test B (April)</td>
<td>12.222</td>
<td>4.790</td>
<td>9</td>
</tr>
<tr>
<td>Test A (May)</td>
<td>13.333</td>
<td>4.416</td>
<td>9</td>
</tr>
</tbody>
</table>

Clearly, the students established higher standard deviation scores compared to the initial test as the study went on suggesting that there may be other factors than the test or intervention which may have influenced their scores. Of course, the size of the sample was only 9 so the group was more vulnerable to being skewed by outlier data. Results for the ANOVA indicated no significant testing effect (Wilks’ $\Lambda; p = .762$).

REPEATED MEASURES ANOVA: PAIRWISE COMPARISONS OF THE ADVANCED GROUP

In order to understand the changes outlined above, pairwise comparisons of each group were conducted. From March to April, the advanced group decreased their paired
difference mean score by -2.636 while their paired difference standard deviation score was 2.942 (see Table 11). It was during this jump in scores from March to April that the advanced group established significance ($p = .50$). However, as discussed prior, this gain in paired difference mean could have been due to the fact that Test B (taken in March) was an inferior test while Test A (taken in April) was the test that all groups did the best on. The study was also relatively novel to the students at this time and they had not yet appeared to tire of testing as the qualitative data suggests many did after the April test (see Efficacy section).

From April to May, the advanced group increased their paired difference mean score 1.909 while their paired difference standard deviation scores also increased to 3.590.

### Table 11. Pairwise Comparisons of the Advanced Class

<table>
<thead>
<tr>
<th>Test</th>
<th>Paired Difference</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test B (March) to Test A (April)</td>
<td>-2.636</td>
<td>2.942</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Test A (April) to Test C (May)</td>
<td>1.909</td>
<td>3.590</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Test B (March) to Test C (May)</td>
<td>-.727</td>
<td>3.771</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>

From March to May, the advanced group decreased their paired difference mean score by -.727, an increase of approximately three-quarters of a point by the end of the study, but their paired standard deviation score was highest at 3.771. A slight paired difference mean gain of -.727 coupled with a high paired difference standard deviation of 3.771 means that advanced students, on average, did not improve much throughout the course of the intervention. However, these results should be interpreted with caution, as there were only 11 students in the advanced group. Therefore, score variations for a few students could have skewed the data set considerably.
REPEATED MEASURES ANOVA: PAIRWISE COMPARISONS OF THE BASIC GROUP

From March to April, the basic group increased their paired difference mean score by .353 while their paired difference standard deviation score was 2.523 (see Table 12). From the first point in data, from March to April, the basic group seemed to decline in performance on their April test compared to March by approximately one-third of a point. They did not experience a surge in performance as the advanced class did. From April to May, the basic group increased their paired difference mean score by .765, declining approximately three-quarters of a point, while the paired difference standard deviation was 4.494. Things appeared to be getting worse for this group as time went by. This might be attributed to the great variance of students in this class consisting of high and low test-takers. However, there were considerably more students in this group, 17, as compared to the advanced group that only had 11. Therefore, a decline in performance for a few should not have had as great of an impact.

Table 12. Pairwise Comparisons of the Basic Class

<table>
<thead>
<tr>
<th>Test</th>
<th>Paired Difference</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Test A (March) to Test C (April)</td>
<td>.353</td>
<td>2.523</td>
</tr>
<tr>
<td>Test C (April) to Test B (May)</td>
<td>.765</td>
<td>4.494</td>
</tr>
<tr>
<td>Test A (March) to Test B (May)</td>
<td>1.118</td>
<td>3.855</td>
</tr>
</tbody>
</table>

Finally, from March to May, the basic group increased their mean score by 1.118 but the standard deviation was slightly lower at 3.855. This suggests that the intervention was actually counterproductive for this group; they actually increased their test scores, on average,
by more than one point from the beginning to end. The decline in performance in the April to May slump that many students seemed to experience could have been from laziness, tiredness, or boredom as evidence by the qualitative data. Many in the basic class did not appear to acclimate to the intervention. This may have been because the needs of the class were so diverse. For example, very advanced test-takers in that group might have been frustrated that we were reviewing concepts and skills that they already knew while less advanced test-takers might have become uninterested in the intervention because they were learning so many strategies simultaneously.

**REPEATED MEASURES ANOVA: PAIRWISE COMPARISONS OF THE BELOW BASIC GROUP**

From March to April, the below basic group initially increased their paired difference mean score by .333, or one-third of a point (see Table 13). However, unlike the other groups who declined in performance from April to May, the below basic group made a gain of 1.111, slightly over one point on average, and while not significant, the continuous decline in paired difference standard deviation scores suggests that, of all of the groups, the intervention had the most positive impact on the below basic test-takers. This may have been because they came in with the lowest skill set and needed many of the strategies (detailed in qualitative section) introduced through the intervention. Initially, the below basic test-takers had a paired difference mean of 5.025. The variance of scores was great but became less over time.

Finally, from March to May, the below basic group decreased their mean score by -.778, approximately gaining three-quarters of a point, on average, by the end of the study and the paired difference standard deviation was slightly lower at 4.295. This suggests that
Table 13. Pairwise Comparisons of the Below Basic Class

<table>
<thead>
<tr>
<th>Test</th>
<th>Paired Difference</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>Test C (March) to Test B (April)</td>
<td>.333</td>
<td>5.025</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Test B (April) to Test A (May)</td>
<td>1.111</td>
<td>4.485</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Test C (March) to Test A (May)</td>
<td>-.778</td>
<td>4.295</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

The intervention did benefit this group slightly since their test mean scores decreased along with their standard deviation scores over time. Their highest decrease in mean could be seen from April to May during the final phase of the intervention unlike with the other groups who made minimal progress during this time. However, although these results might appear promising, once again, the sample size of nine indicates that one or two student scores could have skewed the data considerably.

**Summary of the Quantitative Data**

The results of this study indicate that the student groups in this study did not make a lot of progress over the course of the 3-month study. Advanced students made gains early on from March to April but then decreased in progress from April to May. Overall, they only gained about three-quarters of a point throughout the intervention. The basic test-takers made even less progress. In fact, over the course of the study, they actually lost a little over a point, on average, from March to May. For them, the study seemed especially counter-productive. However, this group was unlike the other groups and had a much larger range of test-takers. The varying levels of test-takers and the different needs might have made some in the basic class lose interest because the intervention may not have suited their needs. Finally, the below basic group made the most gain throughout the study; they increased slightly over a point, on
average, from April to May. However, these results should be interpreted with great caution. Because the class sizes were so low, changes in overall test scores for one or two students could have skewed the appearance of the test scores. Therefore, it is especially important and imperative to analyze the intervention through a qualitative lens to see if a clearer story can emerge.

**INTRODUCTION TO THE QUALITATIVE TEST RESULTS**

The overarching goal of the study was to make the students better test-takers. Instead of presenting a routine test preparation environment in which students took practice tests and went over correct answers in a cursory manner, the goal of this study was to slow down and to look very deeply at a few items per test, particularly the ones that students had the most trouble with.

Focus student think alouds captured the actual thought processes of six test-takers (two advanced, one proficient, one basic, one below basic, and one far below basic) to see how their thought processes changed over the course of the study. They also served to prepare the teacher researcher for potential problems that test-takers might have had on each of the tests since these think aloud results were available for review prior to classroom discussions. Think aloud students recorded their thought processes on every 5th, 10th, 15th, 20th, and 25th item on each of the three Princeton Practice CAHSEE tests.

The goal of the classroom discussions was to allow students to share their thinking processes verbally with the rest of their classmates with the hope that such discussions would impact their decision-making processes not only on future practice tests, but on the actual CAHSEE test itself that they were scheduled to take in March of their sophomore year. Each
time a discussion took place, students first examined the overall test results for half the test that they had taken prior. For example, the advanced students who took Test B in March, looked at their collective results for the first half of the test only, item #s 1-12. After looking at their score distributions, they then voted on three items to discuss in depth. The reason for having students elect to review particular items is that it gave them more ownership and control over their learning. At the end of each discussion, the teacher researcher selected one item from the discussion for the students to reflect on in writing. Students were each asked to complete a journal entry about how they got the answer wrong (or how someone else might have got the answer wrong if they got the answer correct), and how they might approach the test question in the future if they were to see it again.

**Qualitative Results**

During the course of the study, students, in general, showed positive changes in their ability to read and interpret the test question and answers prior to selecting an answer choice. In addition, they showed some improvements in their ability to eliminate answers by using specific evidence from the text. Finally, they showed a trend of revisiting the text more frequently and thoroughly before determining an answer but they continued to struggle with explaining how, precisely, they would revisit the text. The degree and extent to which students expressed their thinking varied from class to class and these specific differences will be explored in the discussion section. In general, a mix of advanced, basic, and below basic students, approximately one-third of the overall students, demonstrated that they understood how to apply the strategies in a cyclical rather than linear process while most students only showed some degree of integration. Below basic students, mostly through the discussion
commentary, expressed improved feelings about their test-taking abilities compared to the other groups (see Table 14). Approximately one-third of the students in all of the groups (advanced, basic, below basic) indicated that laziness or tiredness affected their testing performance.

**Table 14. Continuum of Themed Behaviors Around Test Questions**

<table>
<thead>
<tr>
<th>Verbal Articulation Themes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Reading the question and answers prior to selecting an answer choice:</td>
</tr>
<tr>
<td>Not reading question and any answers</td>
</tr>
<tr>
<td>2. Interpreting the question and answers prior to selecting an answer choice:</td>
</tr>
<tr>
<td>Not interpreting the question and/or answers</td>
</tr>
<tr>
<td>3. Eliminating answers prior to selecting an answer choice:</td>
</tr>
<tr>
<td>Not eliminating any answers and/or eliminating without rationale</td>
</tr>
<tr>
<td>4. Revisiting the text prior to selecting an answer:</td>
</tr>
<tr>
<td>Not revisiting the text through memory and/or referencing</td>
</tr>
<tr>
<td>5. Practicing behaviors # 1-4 concurrently:</td>
</tr>
<tr>
<td>Not practicing or practicing one of the above behaviors concurrently</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Efficacy Themes</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. Feelings about test-taking abilities:</td>
</tr>
<tr>
<td>Expressing feelings of doubt</td>
</tr>
<tr>
<td>Expressing feelings of dislike</td>
</tr>
<tr>
<td>Not expressing feelings of efficacy</td>
</tr>
<tr>
<td>No improvement perceived</td>
</tr>
</tbody>
</table>
Reading the Question and Answer General Results

Results of the study indicated that students' ability to articulate the need to read both the question and all of the answer choices progressed over time. Initially, only one of the six focus students read the question and all of the item choices before making a decision. By the end of the study, all but one focus student did. Through the discussions, 34 out of the 37 students in this study came to recognize that not reading through all of the choices could negatively impact their decision-making process at some point during the study as evidenced by their journal writings. Interestingly, we discovered collectively that reading the question and the accompanying answer choices didn’t always have to be done in a completely linear fashion.

March Findings

At the beginning of the study in March, five of the six focus students repeatedly demonstrated that they did not read through all of the answer choices before selecting an answer. Initially, on the first test in March, five of the focus students did not read through all of the choices from the questions before determining the correct answer. However, the degree to which they did not read through varied. Alejandro, a below basic focus student, was the only one who read through all of the questions and answer choices before selecting an answer. Nathaniel, one of the advanced focus students, read through four of the five the test questions but on one item, he immediately identified an answer without reading through the two of the four answer choice options. Reynaldo, a basic test-taker, read through most of the questions and answers completely, but on one, he went straight for answer choice "A." Bill, a proficient focus student, read through the answer choices on two of the five think aloud
items, but on two of the items, he went immediately to his answer without reading the remaining answer choices.

Rashawndra, an advanced focus student, and Jenaya, a far below basic focus student, demonstrated this skill to even a lesser degree. Neither of them read through any of the answer choices except the one they selected as the correct answer. Rahawndra’s March transcripts were filled with statements like “I think I’m going to go with D because . . .” or “I think it’s A because . . .” On two of the five questions, she began her reading with statements like “Um this one’s kind of tough for me” before she began to read the question. This approach seemed to distract her from completely reading the question and answer choices. Although Jenaya mentioned most of the answer choices, her responses suggested that she did not actually read them. For example, on one item, she said:

Which of the following best summarize the information in the article? I chose A because I think it shows the—It is best to summarize the information in the article. I don’t choose B, I don’t choose C, I don’t choose D because it ain’t best show to summarize the information in the article.

She did not actually appear to read the answer choices she chose to omit.

Two of the focus students, Nathaniel and Reynaldo, offered comments between the reading of the questions and answer choices. On one item, Nathaniel told himself what the text was about. After reading the question, “What makes the preceding sentence ironic?” he said, “Looking at the paragraph, it says that Elizabeth had been climbing for a long time. But, um, she can’t even make it with her friends.” He then continued to read the answer choices and after reading choice A, he added, “I don’t think that’s the right answer because it’s more about Elizabeth and not the brother.” After reading B, he added, “The paragraph is more about Elizabeth than the skill of Maria or Nikki.” He had many intermittent thoughts while he
was reading through the question. Reynaldo, on the other hand, recognized that he needed to read the accompanying reading passage more thoroughly. After reading the question, “According to the passage, the reason that the Cherry Blossom festival is not annually held on a specific date is because the festival is,” he interrupted himself and said, “Oh first, I need to read the text. Let’s just scan it. Don’t take too much time.” Here, he seemed to realize that this was a particular question that required substantial knowledge of the text so he elected to go back and scan before proceeding to the answers.

Prior to the initial phase of discussions following the first test, I saw that, for the most part, the focus students were not reading through all of the answer choices before making a decision. Therefore, during discussions on certain items, I required strict adherence to reading the test question and all of the options before proceeding. For example, I would say, “Who would like to read question 10?” When the student read the question and answers, I would then say, “Now, can anybody say what they would first do after reading the question?” and the discussion would begin. At other times, when a student only read the question and not the accompanying answers, I’d say “And what are the choices?” or “Go ahead and read us the answers please.” I believed that this approach was satisfactory and student journals confirmed this finding. For example, students, when asked, “Why do you think you missed this test question?” or “How do you think someone might have missed this test question?” all student groups tended to answer the questions in a similar way. For example, an advanced student said, “Someone could have missed this question by . . . not reading the question all the way” while a basic student stated, “I missed this test question because I didn’t read the test question carefully,” and a below basic student said, “I think that people could have missed this question because they might not have read all parts of the question.”
However, it wasn’t that simple. As we proceeded onto other items, it occurred to me that reading the question and immediately reading the answer choices could actually confuse the students depending on the question. For example, on the question that asked, “Which source would give the most information for northern lights mythology?” students needed to slow down and really look at what the question was asking. While the article was mainly about the scientific aspects of the northern lights, in this instance, the question complicated the matter by asking the students to seek information on mythology. On such items, it might even add more confusion if students just started to proceed in reading the answers.

I had to realign my approach as I realized that on certain questions, particularly those requiring interpretation, students might not read through all of the answer choices right away, but that they definitely should before making their ultimate decision. For example, on a question that asked, “The selection is best described as A fiction, B biography, C article, or D drama,” I expected them to read the answers immediately. However, sometimes, on more complex questions like one about iguanas that asked, “Which of the following questions could be answered by further research?” I had to help them put the actual test question into their own words. When a student said, “Which question would be answered best if you researched more?” we then proceeded to read the answer choices after we were certain we read and understood the question effectively.

Although I had indicated the importance of reading the question and the answers fully, I now had to step back and recognize that this wasn’t always straightforward, linear process; my approach switched from instructing students to read all of the questions and answers in an immediate, linear fashion to making sure that they read all answer possibilities before arriving at a decision. This meant that they could take time to interpret the question
and clarify their own thinking first so that reading the answer choices wouldn’t make them more confused.

APRIL FINDINGS

During the next phase of testing in April, the focus students made gains in demonstrating their need to read through all of the questions and answer choices before making a decision. They showed this ability primarily during their think alouds. This time, four of the six focus students read the question and answer choices completely before making a decision. Two students, one basic and one far below basic, did not although they read more answers than previously. For example, Reynaldo, a basic student, still managed to read four out of the five questions and answers completely, but on one, he read the question and then simply stated, “Um, I don’t really understand the test question. Ugh, I think I’m going to guess . . . I think it’s B.” Jenaya, a far below basic student, this time, read all of the questions and answer choices on four out of five items, but on one item, she only read the A, B, and C choices, omitting D before making a decision.

Two of the focus students, Nathaniel and Reynaldo, again had intervening thoughts. For example, after reading the passage “The Beach Day,” Nathaniel, an advanced focus student, told himself what the passage was about before he proceeded to read the answers. He said, “Well, I know that this story is about two girls that went to the beach to relax but the spot where they were was too thin so they crossed a creek and found a lighter spot and they set up camp and fell asleep.” Reynaldo, a basic student, on the other hand, told himself what strategy he should apply before reading the answers. For example, on a question that asked, “What makes the preceding sentence ironic?” he immediately said, “Well, let’s use the
process of elimination. It’s better. Hmm, let’s see what wasn’t ironic,” and then he proceeded
to read the answers. Neither of these focus students read the answers in a linear way; rather,
they allowed other thoughts to intervene during the process. Nathaniel refreshed his memory
about what the passage was about while Reynaldo clarified what strategy he would apply.

During the discussions from the second test, students understood the need to read the
question and all of the answer choices more clearly, although when this occurred would
depend on the nature of the question. If a question was straightforward, students would read
the question and then all of the answers. For example, on the question, “What is the author’s
tone? A concerned, B humorous, C desperate, or D critical,” the student would typically read
it straight through. However, if the question was complex and required a great deal of
interpretation, I would intervene and then the process of reading the answers was delayed.
Sometimes, it would lead us to return to the text first. For example, when asked to read a
question about sentence combining from Test A from the advanced group, the following
conversation ensued:

Male Student: Which is the most effective way to combine sentence five and six? Do I have to read all of these [the answers]?
Teacher: Well, let’s not read all of the answers yet. Let’s think about what you would really do. So if you came across this test question,
what would you really do?
Male Student: I would look at sentence five and six.
Teacher: Why would you look at five and six in the text?
Male Student: Because it’s asking me which is the most effective way to combine it.

As this conversation illustrates, teaching students to read through the question and answers
immediately would not serve them well here. Because the nature of this question demanded
that they reconfigure two sentences directly from the text, of course it would serve them best
to first revisit those two sentences instead of reading the lengthy answer sentences in which
three out of four were incorrect. Sometimes, students would have to revisit the text before proceeding to the answers.

While the students did not write about this particular sentence combining question in their journals, this necessity of taking a textual detour before proceeding to the answers echoed in journal responses for other questions. For example, on a question about “Which of the following ideas is supported by details or evidence in the essay [Loch Ness Monster]?” many of the student journals mentioned reading the question and answers in concert with other strategies. For example, an advanced student suggested that someone could have missed the correct answer by “not reading the question thoroughly,” and that someone could now reapproach it “by reread[ing] the essay or reading the questions and answers all of the way . . . .” An average student, on a question related to genre, said, “If someone approached this question again, they should look at the story and the answers and think carefully what to pick.” A below basic student, on a question about finding the best source of information, said, “Next time, I can read better and look at the question then go back to the text and read or scan the text better next time.” Students, in general, were starting to move away from simply reading the questions and answers to other strategies such as eliminating answers or revisiting the text that would increase their chances of not only reading the question and answers, but reading them well. Although reading the answers was crucial in the process, it became apparent that when it occurred depended on the nature of the test question involved.
**MAY FINDINGS**

On the last test, focus students illustrated this process of reading the test questions and answers more consistently. Five of the six focus students read every question and the accompanying answers before proceeding. One student, Jenaya, a far below basic focus student, did not. However, she came close. She read three of the five think aloud questions and answer choices completely, but on two, she only read the question and choices A, B, and C. She simply omitted D both times. Interestingly, none of the focus students added any intermittent thoughts in between reading the question and answer choices like the two who had previously. They simply read the question and answer choices straight through and then offered commentary after they had read all of the choices.

During the classroom discussions from May, students began to demonstrate their own process for what they would do after reading a test question. For example, on the question relating to the reading, “Springtime in Washington D.C.,” the following conversation occurred from the advanced class:

| Teacher: | So 13, can someone tackle that? |
| Male Student: | Which of the following strategies does the author use most frequently to discuss Washington D.C.? |
| Teacher: | Before you even say anything about what the answers are and what they’re not, just tell us what your thought process is. After you read that question, what are you thinking of doing next? |
| Male Student: | Going back [to the text]. |
| Teacher: | And this is somewhat artificial guys because, obviously, we’re looking at the test question a week or two after you took the test, so your memory is not the same as when you’re really taking it, so we have to weigh that in. But because it’s not real familiar, you’d glance back. |

Here, the student immediately read the question without being instructed, and because he only had a vague sense of what the article was about some 2 weeks later, he elected to go
back to the text. Thus, the artificiality of what we were doing emerged as it became clear that 2 weeks after taking the test, the student would have had to go back and read the text before proceeding to the answers. However, if the student had just finished the reading, he might proceed directly to the answers because the reading was so fresh in his mind. Thus, we had just begun to discover the artificial landscape that we were working in. Students began to see that reading the question and accompanying answers was a complex process; sometimes they needed to take more time to understand the question while other times they might need to revisit the text before addressing the answers. In the natural test-taking world, these discussion processes might not mirror precisely what they actually would do while testing. Still, the goal of giving students different steps to take seemed crucial as I knew I could never intervene during that test-taking moment and it would be up to them to make the right moves.

In conclusion, all groups revealed that they understood that they had options prior to completely reading the test question and answers as evidenced by their final journals entries, but they didn’t always state that they would do the same things. For example, on a journal entry that asked students, “Which statement describes Washington D.C.’s setting year-round?” an advanced student said, if he was to see this test question again that he would “go back over the words in the text” while another advanced student said that he would “spend more time trying to differentiate between the two [year-round and annually].”

On another question about the genre of a text from “A Visit to the Doctor’s Office,” a basic student said, “I didn’t read the whole text and I didn’t look closely at all the answers . . . ,” while another basic student said:

I missed the question because I thought that the story was real so I chose answer C, article . . . . I would respond differently by reading the story closer and paying more attention to what the question and answers are saying.
Finally, on a reading about The Loch Ness Monster that asked, “Which of the following ideas is supported by details or evidence in the story?” a below basic student said, “I think that someone could have missed this test question because they didn’t read the passage right or they didn’t read at all.” Still another below basic student said:

If someone were to see this question on a test again, they might approach the question differently by reading the passage a little more closely. Also, if somebody who got it incorrect would like to get it correct they could read each answer more thoroughly.

Four of the 37 students, 2 of whom were focus students, did not mention reading the question and answers at any time in their journal reflections. The reasons for this may have been due to confusion over items or testing fatigue. A below basic focus student said, “I was eliminating answers that had nothing to do with B . . I was stuck,” while an advanced second language student said, “I didn’t get it much. I think most people would get it wrong because they didn’t understand about the words much.” These students appeared to write more about their confusion than how to apply a solution. Still, there were other reasons besides lack of understanding that seemed to influence some students. As an advanced student said, “I missed the question because I didn’t really feel like doing the test.” Finally, an advanced focus student said, “Someone could have gotten it wrong by getting testing fatigue and putting a random answer.” These individuals appeared, like the above students, to be focusing on the problems with the test taker and not what they could do differently in the future. They seemed to misread the writing prompt.
STUDENT SUMMARY RELATIVE TO READING THE TEST QUESTION AND ANSWERS

The students in this study showed strong signs of understanding the value of reading the test question and the accompanying answer choices, although they discovered, along with me, that this process did not necessarily need to be done in a linear fashion and that the actual type of question could influence whether they revisited the question, the text, or both prior to reading though the answer selections. However, two focus students in this study seemed to negate this finding as they tended to offer fewer intervening thoughts in between their reading of the question and answers as time went by. Although nearly all read each question and accompanying answer choices, none of the focus students gave commentary in between their readings by the last test.

In general, the students revealed that they didn’t always need to take the same path. Some sought to revisit the text while others chose to revisit the question and still others opted to revisit the answer choices. There was no one correct way. It depended on the student’s needs and it also depended on the type of test question. I began to recognize that reading through the test questions several weeks after taking the test was somewhat artificial, and that the goal of the discussions was to give them different steps to practice, and that they would be the ones to decide what steps to take while they were testing. Ultimately, their real performance would indeed be an isolated experience.

Interpreting the Question and Answer General Results

Results of the study indicated that students’ ability to articulate the need to interpret the test question and at times, the answer choices, changed slightly over time. Initially, five of
the six focus students interpreted at least one of the five focus test questions before making a decision. However, the degree and extent to which they did this progressed over time. By the end of the study, the same five focus students interpreted the test questions, but now they were interpreting nearly all that they read and not simply one of the five that they were faced with. Through the discussions, 34 out of the 37 students in this study came to recognize that not slowing down to interpret the test question could negatively affect their decision-making process as evidenced by their journal writings. Interestingly, we discovered that beyond misreading the test question, another danger to misinterpreting a test question is to let one’s own prior knowledge outweigh the actual contents of the test question, reading, and answer choices.

**March Findings**

At the beginning of the study in March, five of the six focus students demonstrated interpretative skill for at least one of the five focus questions. However, the degree and extent to which they interpreted the questions varied. Some focus students showed a high degree of interpretation initially while others did not. For example, Nathaniel, an advanced test-taker, offered several comments *between* the reading of the questions and answer choices. On one item, Nathaniel told himself what the text was about. After reading the question, “What makes the preceding sentence ironic?” he said, “Looking at the paragraph, it says that Elizabeth had been climbing for a long time. But, um, she can’t even make it with her friends.” He then continued to read the answer choices and after reading choice A, he added, “I don’t think that’s the right answer because it’s more about Elizabeth and not the brother.”
After reading B, he added, “The paragraph is more about Elizabeth than the skill of Maria or Nikki.” He had many interpretive thoughts while he read through the question.

Others interpreted to a lesser degree. Alejandro, a below basic focus student, responded fairly thoroughly to a test question that asked, “What is the purpose of this article? A to entertain, B to persuade, C to express opinion, or D to inform.” He stated, “I believe it’s D to inform because of the information given. It basically talks about deforestation and they talk about all of these trees being cut down.” Here, he equated that because the article gave information on a topic, the best choice was D, to inform. Reynaldo, a basic test-taker, on a question that asked, “Which of the following BEST summarizes the information in the article?” said, “It is A . . . it summarizes it very nicely.” Here, he recognized the concept of summary and then made his selection. In response to the test question, “By adopted home, the author is referring to ________.” Bill, a proficient test-taker from the advanced class, noted, “I have chosen D Southern California because the adopted home is where they live now.” Here, he was able to interpret adopted home in his mind as a place where the Iowans live now, the place they have relocated to.

The student who did not show any signs of interpretation was Jenaya, a far below basic test-taker. She did not read any of the questions during the March test. However, Rashawndra, an advanced test-taker, who only read one of the five test questions during her March test, still showed a moderate amount of interpretive skill. She stated, “I think it’s A because it basically sums up the whole story of what just happened and like it says, which statement BEST describes what happens in the story and A is what happened in the story.”

Prior to the initial phase of discussions following the first test, I saw that, for the most part, the focus students were interpreting randomly and not thoroughly. None of the focus
students demonstrated a high degree of interpretive skill on several of the five focus questions. Therefore, during classroom discussions on certain items, I recognized that I needed to encourage the students to slow down and absorb the test question before proceeding to any other strategies. For example, during a discussion on "The Incredible Northern Lights," that asked "Based on the essay, which of the following would be the BEST source of information to use to learn more about northern lights mythology?" many students, more than half of all of the groups, missed this test question. Many did not see the word *mythology* and so they chose an answer that was scientific, and therefore, incorrect.

During the discussion, I stressed the idea of stopping and looking carefully at the test question, to "put it into your own words." When I asked advanced students to put that test question in their own words, I got responses like, "Use the question to base it on yourself," "Read the whole question and make sure you know—like have some past knowledge about the question," or "Use the vocabulary in both the question and the answers to figure out which one works best." The students did not seem to understand how to put the question on northern lights mythology into their own words so I had to model it. I stated:

If you’re putting it in your own words, just pretend that you’re that person who is trying to find out the information. So you want to learn more about northern lights mythology, which source is best? What happens is the question is worded so bizarre that by the time you get to the end of the question, you’re thinking about northern lights and you’re not necessarily thinking about mythology.

At this point, I felt I needed to stress that it’s crucial to understand what a test question is asking before you proceed. I added:

Don’t make your decision on your answers unless you’ve really wrestled through that question and you know what it’s asking. In this case, it’s asking you to get a good resource for northern lights mythology, not scientific information . . . .
At this point, students could recognize how misinterpreting the test question could set them up for failure. Now that they could clearly see what the test question was asking, they could easily eliminate other choices. Students collectively recognized that they sometimes missed answers because they misread the test question, possibly because they did not take the time to paraphrase the test question before seeking a correct answer. At times, I had to draw this out of them. After a below basic student read, “Which of the following sentences of the article helps to describe the year-round setting of Washington D.C.?” the following conversation occurred:

Teacher: ... So what should you do at this point to make sure you’re heading in the right direction?
Male Student: Read it [the question] over.
Teacher: ... If you tried to put it in your own words, what’s the question asking you?
Male Student: Which ones the right one.
Teacher: But for what?
Male Student: The question.
Teacher: What part of that question, if you put it in your own words, [what] is it asking? What’s it asking you to pick?
Male Student: Describe the year-round setting of Washington D.C.
Teacher: So what’s a key word that you’re trying to pick?
Male Student: Describe.
Teacher: Yes, and what kind of setting?
Male Student: In Washington D.C.
Teacher: But what kind of setting?
Male Student: Year round.
Teacher: So that would be?
Male Student: All year.
Teacher: And all?
Male Student: Seasons.
Teacher: So that’s a really key part of the question . . . .

Students were beginning to see that before they went through the trouble of eliminating answers, they had to really be clear about what the question was asking.
The idea of not understanding the question was a problem on other items as evidenced by student journals. As one advanced student wrote in her journal reflection on another question related to “The Iowa Coast,” “Next time, I will read the question repeatedly until I properly understand what the question is asking.” Here, this student recognized the need to wrestle with what the question is asking before proceeding. On a question that asked, “Which of the following details is supported by details or evidence in the essay?” from the reading “The Loch Ness Monster,” a below basic student said, “If someone comes by this question again in the passage or in a test, they can get the question right if they take there time and reread the question.” This student recognized the need to slow down and reread in order to absorb what the question was asking before proceeding.

APRIL FINDINGS

During the next phase of testing in April, the focus students made gains in interpreting the test question before making a decision. They showed this ability primarily during their think alouds. This time, all six of the focus students showed interpretive ability on at least one of the five questions. Several made gains by interpreting multiple questions this time around.

For example, Jenaya, a far below basic test-taker, was now showing signs of interpreting. After reading the question “What makes the preceding sentence ironic?” about Elizabeth and the mountain climb, she stated, “The correct answer for question five is the fact that although Elizabeth was an experienced mountain climber, she was the only one who was having the most trouble climbing the mountains.” She was beginning to show signs of interpreting the question by linking her understanding of the text back to the question.
However, the degree to which she did this was somewhat implicit, while other focus students showed stronger signs of explicitly interpreting the text. For example, Reynaldo, a basic test-taker, seemed to interpret more clearly, actually recognizing key words in the question. He stated, “Hmm. Let’s use process of elimination. Let’s see what was ironic.” Before beginning to eliminate, he knew that he was looking for an ironic statement.

Still others showed a higher degree of interpretation and they did this multiple times on several questions, unlike in March. For example, on a question that said, “Instead of following a strict vegetarian diet, the author’s second article encourages the reader to . . . ,” Alejandro, a below basic test taker, said, “I don’t think A avoid eating vegetables and eat meat only. It doesn’t sound right. It’s saying strict vegetarian diet so it doesn’t go with it.” Here, Alejandro acknowledged the word strict and therefore ruled out a possibility because the answer choice did not match. Thus, he was beginning to hone in on key words in the test question. He continued to do this on other items as evidenced by the following rather in depth response from the reading, “Springtime in Washington,” on a test question that asked,

“According to the passage, the reason that the Cherry Blossom festival is not annually held on a specific date is because . . .”:

I don’t think it’s um A. The festival is held on the date that the mayor of Tokyo comes to Washington. I don’t think that’s it because it says right here, in the text, that the Cherry Blossom festival occurs just as a sea of sweet-smelling, pink flowers brightens the city. Although the other Mother Nature knows exactly that, um, clearly it’s C because it says that it’s impossible to predict exactly when spring’s warmer weather will arrive. They don’t know when it arrives so when it does, that’s when they have the festival. B, the parade organizers like to hold the festival when the most tourists arrive in town. That’s a good one but it doesn’t make sense with what it says about the spring. D, the president changes the date of the Cherry Blossom festival every year. That doesn’t say that so I think it’s C.
Alejandro demonstrated a much higher level of thinking than previously through this response. He obviously approached the test question through the lens of looking for a reason why the date of the festival changes every year. He interpreted the question effectively as evidenced by his exhaustion of all possibilities. He even continued to eliminate other answers after he already identified the correct choice.

Others, like Bill, a proficient test taker, showed great improvement in interpretive ability. On a question that asked, “Which statement BEST describes what happens in the story?” he added:

I’m choosing answer A because answer B only refers to the first part of the story. Answer C does not refer to another part of the story and answer D only refers to another part of the story but not the entire story so answer A is the best choice.

Like Alejandro, Bill was beginning to hone in on the key concepts in the question. Clearly, he recognized that the question asked him about the statement that best describes what happens in the story. He double-checked all other responses and realized that although many of the other responses were true, they did not encompass the whole story. Therefore, he interpreted that he needed to find the best answer, the one that encompassed the essence of the story.

During the discussions from the second test, students demonstrated that they understood the need to interpret the question and all of the answer choices more clearly, although the degree to which it was done would depend on the nature of the question. For example, if a question were straightforward, like “This selection is BEST described as A fiction, B biography, C article, or D drama,” then the students would not take much time to interpret the question. Rather, they would be more focused on which genre type fit the story best. However, if the question was complex, like “The following are references for sources of information about food. In which of them would you be most likely to find information on
how to balance a vegetarian diet?” the level of interpretation was greater. Because this was
more of a layered question, the job of interpreting it was more difficult as the following
conversation with a below basic class illustrates:

Teacher: So in your own words, what are they asking you in this test
question, if you have to put it in your own words?
Female Student: Like which choice you would find the most information on the
vegetarian diet—like which one is the most likely to be chosen.
Teacher: And what’s the key word in the question?
Female Student: Most.
Teacher: Most likely to find information on what?
Female Student: Vegetarian.
Teacher: Not just vegetarian. What does it say?
Female Student: How to balance.
Teacher: Yes, how to balance is a key word there. So again, when you’re
looking at a test question and you’re trying to put it in your own
words, you’ve got to wrestle with what it’s trying to ask.

Clearly the students in this class struggled to articulate what the test question was asking. But
after several probing questions, they were able to pull out what they needed to look for before
addressing the answer choices. Not only was the term “most” important, but so was
“vegetarian” and the phrase “how to balance.” Thus, the students were learning to read with
new eyes as they began to see that they had to be very clear about what the question asked
and what they were looking for before they proceeded.

This similar theme of establishing new eyes was also apparent in their journal
responses. For example, an advanced student noted that, “I think people could have missed
because of not understanding the difference between annually and year-round,” while a basic
student said, “I got the answer wrong because I didn’t read the whole question and see [the
word] myth.” Still another, below basic student said, “I didn’t really pay attention when it
said it was a story. [In the future], I would try to pay more attention next time and find more
clues.” Thus, the students were beginning to identify key words as clues. The advanced student realized that understanding the difference between two words, annually and year-round, was the key to interpreting the test question while the basic student realized that because she did not “see” the word myth, she misinterpreted the test question. Finally, the below basic student realized that there were clues already there, that the introduction to the reading did in fact, refer to a story, but that he did not pay attention to this information. Thus, all of the groups showed some new level of awareness about how to interpret, recognizing that they had to zero in on the key words and really absorb what they meant.

**MAY FINDINGS**

On the last test, most focus students illustrated this process of interpreting the test questions and answers more fully. Five of the six of the focus students showed some improved ability to interpret the test questions by the end of the study and the degree to which they were noticing key words in the test question was becoming more evident. For example, on a test question that asked, “Which of the following BEST summarizes the information in the article?” Bill, a proficient test-taker in an advanced class, said, “I am choosing A because looking back on the text, it makes the most sense. And that is my choice.” He then added, “On second thought, I am going with C because it actually makes more sense than A.” Although Bill did not clearly articulate why he was eliminating answers, he was clear in his mind that he was looking for the statement that best summarizes the article. Since he interpreted the test question clearly, he had a better chance of getting it right. And he changed his mind because as he exhausted the other possibilities, he found what he considered to be a better choice.
Reynaldo, a basic test-taker, had also developed strategies for interpreting test questions. He responded to a question asking “What does the word integral mean in the following sentence?” by stating:

I think the answer is D because using plug and check if you plug in the word, if you replace the word with the four options and you see which one makes sense, essential sounds better than all of them in all.

Here, Reynaldo was showing that through using a plug and check method of substitution and by exhausting all possibilities, he was increasing his chances of getting the question right. Rashawndra, an advanced test-taker, responded the following way to a test question that asked, “Which of the following questions could be BEST answered by further research?” She said, “Um, I kinda think all of these could be answered by further research. But the key word was best because it’s all bold and capitalized.” Thus, she determined that the test-maker was looking for the response that could yield the most research.

Jenaya, a far below basic test-taker, was the only focus student who didn’t show additional signs of being able to interpret the test questions. She didn’t seem to pay any attention to the clues like MOST and BEST in the answer selections. Her think aloud transcripts from May still revealed that she was making haphazard decisions without sound rationale. However, her May journal responses indicated that she was using a process of crossing answers out. As she stated, “...I got the rong answer. I didn’t crosse out the answers. I will respond by reading the questions and rereading the article. I will cross out and then get the correct answer.” Perhaps the language barrier that she faced affected her ability to put the test question into her own words.

During the classroom discussions from May, students began to demonstrate their own process for interpreting test questions. They needed me less. For example, on the question
relating to the reading “Springtime in Washington D.C.,” the following conversation ensued from a basic class:

Teacher: So look back at the question [Which of the following sentences of the article helps to describe the year-round setting in Washington D.C.]. What’s the question asking you for in your own words?

Male Student: What does Washington D.C. look like year-round?

Teacher: Okay good. Notice how he put that in his own words. What does Washington D.C. look like year round? Now if he said ‘What does Washington D.C. look like?’ that would not be so good, how come? What does the question ask?

Male Student: Year-round.

Teacher: Yeah, year-round is really a key word. So many times when people miss these questions, they don’t see like a certain phrase in the question.

Here, the student made the jump and was able to interpret the question in his own words without omitting the important parts. He was still able to maintain the essence of the question but he had put it into his own language. Thus, students began to discover that they had to own the test question by putting it in their own language. This was a very important preliminary step prior to exercising any other strategies.

Journal entries supported this idea of interpreting, in addition to the dangers of superimposing one’s own prior knowledge into the decision-making process. For example, on a question related to a vegetarian diet, an advanced student said, “I think most people got it wrong because they just read and base on there mind, what they think and they forget to look back at the question.” Here, this student noticed the dangers of allowing one’s mind to superimpose information that outweighs what is in the actual test question, a form of misreading. Another student from a below basic class discovered this similar idea through a different question. On the question, “What is the purpose of this article? A to entertain, B to persuade, C to express opinion, or D to inform,” she stated:
I got this wrong because I misread it. I thought it was an opinion so I got it wrong. Next time I see this question again, I won't put my opinion on it so that way I'll be able to know the question.

Here, she recognized that she put too much of her own opinion into her reading of it and she discovered that she should limit this opinion and look more closely at what the test question is actually asking.

Finally, an average student responded the following way regarding the question, “Based on the second article, which of these statements is true?” She said:

Someone could’ve missed this question because they could’ve put their thought about their opinion on diets into their answer . . . . If someone was to see this question again, they would probably go back over the article. The person might also think about what the more correct answer is without using their opinion.

Now students were not only aware of the dangers of not only of not seeing the key words in test questions, but of actually adding prior knowledge that outweighed information from the text, the answers, or the question.

In conclusion, all groups revealed that they understood that interpreting the test question was important. They recognized the importance of paying attention to key words, sometimes rereading the test question for greater clarity, and finally of putting the test question into their own words before proceeding to an answer. As one advanced student said, “. . . if I were to see this test question again, I would probably examine each clue word they give me within the question and take my time to think about it.” Thus, the importance of looking for clues in the question became apparent, long before students took on the arduous task of eliminating the answers and revisiting the text.

Three students of the 37, one of which was a focus student, did not mention interpreting the question and answers at any time in their journal reflections. The reason for
this may have been a desire not to write in their journals or to write vague responses. One basic student, said, “I would go back to read the question and the text again before I answer it,” but he did not talk about interpreting. Instead, he said things like, “I guessed on it. I do not know what the question means,” or “I missed the test question because I was not trying and was not reading the question clearly.” He seemed to really struggle with applying himself to the intervention. Another below basic student said, “If I were to see this question again, I would focus better because I was just skimming and scanning fast in a hurry.” This student seemed to feel that because the intervention did not count toward his grade, he wasn’t going to apply himself fully. Finally, a below basic focus student said, “Next time, I do a question similar to this one, I am going to take evidence and put it together to find the author’s main purpose.” However, he did not specify where he would get the evidence from: Would it be from the test question, the answer choices, and/or the text? His answers were vague and he did not elaborate. These students may have suffered some degree of testing fatigue. They were not showing the ability to work past their confusion.

**Student Summary Relative to Interpreting the Test Question and Answers**

The students in this study showed strong signs of realizing the complexities involved interpreting the test question and the critical value of understanding what the test question means before proceeding to an answer. They discovered, along with me, that one of the grave threats to interpreting clearly is by bringing prior knowledge to the question to the point that it superimposes, or even outweighs, what the test question is actually asking. Although five of the six students progressed in their interpretive skill over time, one, a below basic test-
taker, did not write clear explanations in his journal while another, a far below basic test-
taker, struggled not only to interpret but to read the test questions.

When it came to interpreting the test question, students were more likely to make more errors than simply from reading. They had the added challenge of honing in on key words and putting the test question into their own language. Interpreting was certainly a skill that demanded a higher level of thinking than simply reading. Through the discussions, students were able to see the critical need to know what the test question means prior to proceeding. Through their journals, many mentioned the importance of interpreting clearly, particularly the value of recognizing key words, and the dangers of bringing too much prior knowledge to the table.

**Eliminating Answers Prior to Selecting an Answer Choice General Results**

Results of the study indicated that students’ ability to articulate the need to eliminate the answer choices changed dramatically over time. Initially, none of the six focus students eliminated *all* of the alternative answer choices before making a decision. However, the degree and extent to which they did this progressed. By the end of the study, three of the focus students eliminated the answers and now they were eliminating nearly all of the choices that they encountered prior to making a decision, and they were doing this with sound rationale. Through the discussions, 36 of the 37 students in this study came to recognize that reading through and eliminating all other answer choices prior to selecting an answer was the best way to increase their odds of getting a question correct at some point during the study as evidenced by their journal writings. Interestingly, students discovered that when faced with
two choices, it was in their best interest to select the answer choice that encompassed the biggest, broadest picture.

**MARCH FINDINGS**

At the beginning of the study in March, four of the six focus students demonstrated some degree of ability to eliminate answers with rationale. However, none of these focus students exhausted all of the alternative choices before selecting an answer. For example, Nathaniel, an advanced focus student, often was very logical in explaining why he got rid of certain answers, but he didn’t exhaust all possibilities. On a question about iguanas that asked, “Which of the following questions could be BEST answered by further research?” he said:

> It’s B I think because in the fourth paragraph, it talks about for the metabolism or body chemistry to function properly—properly, they need tropical sunshine and I doubt that any artificial light is ever truly an adequate substitute. B is also, I also think it’s B because the writer says I doubt so the writer uses his own opinions, in with the facts to make this article more enjoyable, readable, and more informational.

Even though he provided a strong rationale, he didn’t eliminate the other choices. He simply found the one he thought was correct, gave good reasoning, and then proceeded onto the next test question.

The remaining focus students, one proficient and one far below basic, did not give any rationale for their selections. Bill said things like, “And the answer I’m choosing is C because it fits and answers the question perfectly” while Jenaya said, “I think A because it means horde. I didn’t choose B because it wasn’t correct. I didn’t choose C because it wasn’t correct. I didn’t choose D because it was not best for the answer.” Thus, early in the study, it became apparent to me that focus students, regardless of their testing level, were not
exhausting all answer possibilities before arriving at a decision. Therefore, during the discussions, I required students to eliminate all answer choices with appropriate rationale before deciding.

During the discussions, I stressed the idea of exhausting all answer possibilities. As I did this, students collectively struggled, at times, to get rid of alternative answer choices as the following basic class conversation illustrates from the excerpt “Writing a Short Story”:

Teacher: D? A well developed plot, first, you must think of a good story. Why doesn’t it make sense?
Male Student: It’s weird.
Teacher: Any other ideas?
Female Student: It’s backwards.
Teacher: Good, so the order of the sentence sounds wrong, good. So now we’re down to A, B, and C. What can we do here?
Male Student: Eliminate B.
Teacher: First, you must think of a good story with a well developed plot. Why?
Male Student: Because—never mind.
Teacher: Why did you say “never mind”?
Male Student: I don’t know.

The male student in the above example was not able to articulate why he wanted to eliminate B. He simply said, “Because—never mind.” At this point, someone else in the class helped him to provide a rationale. And this theme, of one student helping another to articulate, could be seen in all classes. However, sometimes a student would impulsively want to choose an answer without exhausting all possibilities and would need redirection. This is demonstrated in the following conversation with a below basic class on a test question from “Welcome To Our Company/Cake Director” that said, “The main idea of the passage concerns __________.”

Teacher: D, the best way the receptionist can help the cake decorator. Why did you say it’s not D?
Male Student: Because I said so.
Teacher: Well, keep thinking about it.
Male Student: It’s not D because they’re not talking about a receptionist.

At this point, one student was not able to provide rationale but another student assisted. As the conversation continued, students began to see the importance of selecting the biggest, broadest answer:

Teacher: ... Now, we’re down to A, B, and C. What can you get rid of next?
Male Student: B.
Teacher: Can you explain why please?
Male Student: Because it doesn’t explain like the best way to decorate a cake. It explains what you need to do to decorate a cake—like it explains what you need to get the job and what you need to do for the job when you get it.
Teacher: Very good. So he’s saying here that it talks about the whole job. But a specific part of the job would be what?
Male Student: Baking the cake.
Teacher: Or decorating the cake.
Male Student: That’s what I meant.
Teacher: This one—this [example] is really good because you guys are going to see this come up over and over again. This one is kind of too small of an answer because it’s just talking about the best way to decorate a cake. So the job description is bigger than that. Do you see that?
Male Student: Yeah.

After I pushed students to provide rationales and once they narrowed their answers to two choices, I introduced the idea of selecting the bigger picture over the smaller picture answer as the following example illustrates when this advanced class got it down to C and D on the test question that asked, “Which of the following ideas is supported by details or evidence in the essay?”

Teacher: ... So is C a correct statement?
Male Student: No.
Male Student: No.
Male Student: Yes, it is.
Male Student: Yeah.
People during the Middle Ages had wild explanations for everything that they saw in nature. That’s in the second paragraph that you read. So, that is correct information from the article, so why isn’t it right?

There’s a more correct answer. Yes. And probably one of the most difficult things about these tests are as soon as you see a correct answer, you’re like, that’s the answer. But then, you got to really, really, really make sure that it’s the most correct answer. So if you’re taking these tests and you’re just trying to recognize the correct answer, you can end up missing a lot, because there’s going to be more than one correct answer.

Students, at this point, could see the value in eliminating all answer possibilities in order to prevent them from getting the answer wrong. And when they had difficulty deciding between two choices, they were now introduced to a conceptual strategy of selecting the bigger picture instead of the smaller picture to fall back on. This concept became very apparent in their journal reflections.

One advanced student noted:

If I were to see this question again on another test I would read the question and look into what they are really looking for . . . . Lastly, I would pick the most BEST answer, or the answer that has a BIGGER PICTURE.

This student recognized, through the discussions, that the answer that encompasses the broadest, most general scope would be most likely to be the correct one. A basic student associated the answer with the greatest amount of evidence. She noted, “I got this answer wrong because I got confused and because I didn’t think I had a lot of evidence . . . . I would go back and read [the answers] carefully and see which one has the most evidence.” Here, the student explained that the strategy of reinforcing an answer choice with evidence would increase her odds of getting the test question correct. Finally, a below basic student incorporated the idea of eliminating all of the choices prior to selecting an answer. She said:
I think that if I eliminated all the completely wrong answers, that would have helped a lot. I think I would get it right if I took the test again because I would definitely use the process of elimination to get it right.

This student seemed entirely unaware of the strategy of getting rid of answers before selecting an answer until our discussion.

Thus, among the three groups, students identified myriad strategies for identifying answer choices. Some would suggest narrowing the choices down to two and then selecting the answer that encompasses the bigger picture while others would also revisit the test question before deciding. Some also mentioned the importance of going back to the text when getting it down to two choices before confirming. The important thing was that students were gaining awareness that they could apply more than a single strategy to the task.

APRIL FINDINGS

During the next phase of testing in April, the focus students made gains in eliminating answer choices before deciding on an answer. However, these gains varied among students. One advanced student, Nathaniel, eliminated all answer choices within items with sound rationale before deciding on an answer. But he was the only one. Four of the six focus students showed a pattern of selecting an answer with strong rationale but not always eliminating the remaining choices with rationale. For example, Reynaldo, a basic test-taker, responded the following way on a test question that asked, “What makes the preceding sentence ironic?”

Well, let’s use the process of elimination. It’s better. Hmm, let’s see what was ironic. C, A doesn’t make sense. B or D. Well, yeah because C—the fact, the part ironic is that she is a veteran. She is experienced if she’s a veteran so she’s experienced at climbing mountains although only three hours into the climb, just three hours, not even a long time into the climb, she just gave up and said it was too difficult. The ironic part is that she’s experienced and she’s having trouble
climbing the mountain. So, it would be C, the fact that although Elizabeth was an experienced climber, she was the one who was having the most trouble climbing the mountain. So C.

Here, Reynaldo put a lot of thought into his selection. He knew he was looking for an ironic statement, and he realized that because Elizabeth was a veteran climber, it was ironic that she gave up so soon. However, Reynaldo did not explain clearly why C or A did not make sense, and why D was an inferior answer to C. All of these four focus students, at times, showed patterns of impulsively eliminating without a rationale but they had improved dramatically with respect to providing a rationale for their choices. However, Jenaya, a far below basic focus student, struggled to provide any rationale for her answer choices although she was now reading through the test question and most of the answer choices unlike before. Jenaya’s struggles may have had to do more with her comprehension. In the above example, it is clear that Reynaldo understood key terms like ironic and veteran. Without knowing those words, the above question would be very difficult to get correct.

During the discussions from the second test, students understood the idea of the need to eliminate answers much more clearly and they began to see that getting the choices down to two was a way to make confronting the test question much more manageable. However, once they got it down to two, they struggled with what to do next as the following conversation from a below basic class illustrates about iguanas. The test question asked, "Which of the following most accurately indicates the author’s attitude towards iguanas?"

Teacher: ... So if we get rid of C, what else could we get rid of?
Male Student: B [They are slow and stupid and unworthy of study].
Teacher: Why?
Male Student: Because they’re not stupid animals.
Male Student: And it’s an opinion.
Male Student: And it’s also an opinion like James said. And nothing is unworthy of seriously studying, if it’s an animal.
Teacher: ... It's between A and C—we better look back at the text. ... What's he saying there? [They are poorly adapted to their environment and unlikely to survive].

Male Student: That they're not good pets! But that's not true. That's still sort of an opinion because my cousin had an iguana and he took care of it and nothing—

Teacher: I know people who have had them as well, and they can be good pets if they're taken care of right, but what is the question asking you?

Male Student: What's the author's attitude towards iguanas?

Teacher: The author's attitude.

Male Student: But he's just saying a statistic though.

Teacher: But even if you disagree with the author and maybe you think that iguanas make fine pets, do you understand why it's still C because you're trying to figure out what who's thinking?

Male Student: You.

Teacher: It's not asking about you. It's not asking which of the following do you think—what's your attitude towards iguanas? It doesn't say that in the question so you can't give your attitude.

Here, the students struggled with superimposing their own ideas onto the test question and this gravely affected their decision-making process. Such confusion often became apparent with all three groups on test questions that asked for the author's perspective. This idea of having to learn to differentiate between their own opinion and the author's was apparent in their journals. For example, regarding a test question about author's tone, an advanced student said:

Based on our discussion, how someone could have missed the test question is they get their mind blocked by there own opinion because the author really didn't say he was on a certain side ... Based on our discussion, how someone might approach this test question differently if they were to see it again is instead of assuming which side the author is on, be 100% sure by narrowing the answers after reading the question and then go back to the text to refresh your mind of what tone he is most likely using. Then, you would look at the two answers and remove the answer that has the ability to have any flaws attached to it. For example, since the author never stated which side he is on, he can't be critical about the topic but instead he is concerned because it says “worrisome.”
Here, this student showed that she had a multitude of strategies to get the answer correct and that to be 100% sure, she must look back to the text and search for key words. In this case, the word “worrisome” convinced her of the author’s tone.

Through their journals, other students mentioned the importance of confirming answer choices through key words, although their responses were not near as elaborate. A basic student remarked, “The way someone might have approached this differently next time is reading certain key words in the question to help eliminate wrong answers.” Still, a below basic student remarked, “If you read and you don’t find the correct one, try to eliminate the unmatching answers, such as things that aren’t related” while another below basic student added, “I would magnify each and every word so I wouldn’t get it wrong.” Thus, through the classroom discussions, students were beginning to see that paying attention to all of the words, whether it was “tone,” or “ironic” in the test question, or “veteran” or “worrisome” in the text; these clues could lead one closer to the correct answer.

**MAY FINDINGS**

On the last test, most focus students illustrated this process of eliminating the answers more fully. Three of the five focus students showed a more thorough rationale for selecting their answers, but they still sometimes neglected to clearly explain why they would eliminate all of the other choices. For example, regarding the question involving Elizabeth and the mountain climb, Rashwandra, an advanced test-taker, said:

Um, I’m going back to the text . . . this is hard. It’s not B because Maria and Nikki were not veteran climbers ‘cause it says in the first paragraph . . . that only Elizabeth was the good climber. So yeah, it’s not B. Um, I think it’s C because it said Elizabeth was a veteran climber and that she gave up three hours into it. And the other ones kept going . . . Ironic, I guess is when something contradicts something. And she’s kind of contradicting the fact that she’s a veteran by not
climbing the mountain. And so I guess the word ironic is the key for me. So, yeah. And I don’t think the other answers sound right so I’m just going to go with C.

Although Rashawndra did not fully explain why the other answers were correct and she only said that they didn’t “sound right,” she did an excellent job of providing rationale. And on all other test questions, she eliminated every possible answer. This was the only question that she did not eliminate completely so she showed much sharper skill as time went on compared to her first think aloud transcript.

Other students showed moderate abilities to provide a rationale and leaned more on the process of elimination to help them decide. For example, when asked, “Which of the following questions could BEST answered by further research?” Alejandro, a below basic test-taker, said:

I think it’s B, how does sunshine affect an iguana’s metabolism because it’s the most specific and it is the best answer for further research because the other ones don’t really make sense. Do iguanas enjoy swimming? That’s not it. Do captive iguanas want to return to the wild? Or what do insects taste like to iguanas? The answer’s B.

All focus students, with the exception of Jenaya, provided sound rationale for the answer choice that they selected although they didn’t always go through the process of eliminating every single possible answer.

It’s possible that the type of test questions students were asked influenced how in depth they would go before selecting an answer. For example, regarding the question involving “What does horde mean in the following sentence?” Bill said, “I am thinking it is D [crowd] because I have heard the word before and it usually means a group of people so I am going with D because I have prior knowledge about it.” Thus, like the student responses during discussions and journals, sometimes the nature of the test question and what students
already knew could affect how they approached certain test questions. If they already knew what a word meant or they felt absolutely confident about their answer through evidence from the text, they would not always eliminate all answers before making a decision.

During the classroom discussions from May, students showed a stronger degree of rationale for why they would eliminate certain answer choices as the following advanced discussion illustrates from the passage, “The Pros and Cons of a Vegetarian Diet.”

Teacher: So, what’s the question asking?
Male Student: What is the author trying to accomplish?
Teacher: Yeah, you have to forget what you think and ask yourself, “What is the author trying to do?” This is hard because we have to separate from being the author. Okay, what can we get rid of?
Male Student: C.
Teacher: The author is asserting that people who become vegetarians get sick less often than meat eaters. Why would you get rid of that?
Male Student: Because it doesn’t really say anything about getting sick, except in that comparison when he says, “tying a dirty sock around your neck will cure a sore throat.”
Teacher: Yeah, there’s no real solid evidence that he goes into depth on somebody getting sick, it’s just a small mention. So that answer is probably too narrow. Good.
Male Student: I’d say A, [the author is trying to make the reader laugh by painting a vivid mental picture by using imagery].
Teacher: Okay, good. You think this is way off base. He’s not being funny—his tone is not funny and you’re right. So we’re down to B and D [The author is mocking the idea that vegetarians have lower cholesterol than meat eaters or The author is trying to make the idea of switching to a vegetarian diet to lower the risk of cardiovascular disease seem slightly foolish]. Now only one person in the class got this right, is that right? . . . Is there anyone who can explain why one is better than the other? . . .
Male Student: In D, it says cardiovascular disease but it doesn’t say it in the paragraph.
Male Student: Yes it does.
Teacher: That’s all right. There is a clue in one of those that should tell you that it’s not the right one.
Male Student: One thing that looks a little different is the words “slightly foolish” and “mocking.”
Teacher: You got it. So, slightly foolish means slightly foolish. What does slightly mean?
Male Student: Very little.
Teacher: So basically, if you read this with foolish, the author is trying to make the idea of switching to a vegetarian diet to lower the risk of cardiovascular disease seem foolish. That might be correct, but what do you have?
Male Student: Slightly.
Teacher: Slightly foolish. It’s not strong enough. The one here—this word is stronger. What is this word?
Male Student: Mocking.
Teacher: What does that mean?
Male Student: Making fun of.
Teacher: So the author is mocking the idea that vegetarians have lower cholesterol . . . .

Here, the student had to differentiate between two similar terms, “slightly foolish” and “mocking,” and it seemed to be the adjective “slightly” that actually made the difference. Thus, during these discussions, often choosing the correct answer came down to one key word that created a different shade of meaning. They were beginning to see that every single word of each test question and answer counted and that such words were strategically placed by the test makers. They were becoming aware that differentiating the key words was essential to getting the answer correct and that slight changes in the way phrases were worded could mean the difference between them selecting a correct answer, or the most correct answer.

Journal entries confirmed this finding in all classes. A basic student remarked, “I would do better next time by narrowing down all the answers, paying more attention, and be sure that I don’t miss anything and that I know also know what some words mean.” A below basic student said:

I do understand how somebody got this question wrong. They might have confused “to persuade” with “to express opinion.” It does sound like the author is expressing opinion but he is really informing the tester. If someone got this wrong, the next time they see it, they could use process of elimination to eliminate all the answers that are absolutely wrong.
Finally, an advanced student said:

If I was to see number 11 on a test again and if I was unsure of what the answer was, I would look back to the full text, read it more fully . . . . Then I would try to connect the questions with the text and see which of the options is more true and is actually backed up with evidence from the passage. Because I assumed and didn’t see from more of the author’s point of view, I got the question incorrect.

Students, collectively, determined that individual words meant a lot and that when in doubt, it was best to reexamine the evidence with a critical eye.

In conclusion, all groups revealed that they understood that eliminating the answers was not only important, but a critical process. They recognized the danger of falling prey to relying on their own knowledge and information too much. They learned to pay more attention to what individual key words said, and they were beginning to see these words as clues that could help them to battle the testing landscape. They began to see themselves as capable test-takers with an arsenal of skills, skills that they had built up through interacting and learning from each other.

One of the 37 students did not mention eliminating the answers. However, this student did not mention any of the other strategies that the students had mastered. He left several journal entries blank and did not participate in the class discussions. He had trouble paying attention and he stated in his journal that “I don’t know what the question means,” or “I don’t care,” or “I did not learn anything because I was asleep.” He may have suffered from attention deficit disorder or he may have had other critical thoughts on his mind besides testing.
STUDENT SUMMARY RELATIVE TO ELIMINATING THE ANSWERS

The students in this study showed strong signs of learning to eliminate the answer choices before proceeding to an answer. They discovered that one of the grave threats to eliminating incorrectly can happen by not noticing key words that serve as clues. They discovered, like me, that superimposing their own ideas onto the text can be very dangerous, especially on questions that ask for the author’s point of view.

Revisiting the Text General Results

Results of the study indicated that students’ ability to articulate the need to revisit the text changed over time but their ability to cite specific evidence within the text did not change much. They tended to just rely on memory. Initially, one of the six focus students revisited the text prior to selecting an answer choice while four of the six focus students only provided details of what they had remembered reading. By the end of the study, only one focus student revisited prior to making decisions through citing specific portions of text for all possible answer choices, not just the answer that he selected. Three of the focus students mainly cited specific evidence when deciding on an answer choice only while the remaining two only cited evidence from memory. Through the discussions, 36 of the 37 students in this study came to recognize that revisiting the text was a crucial part of increasing their odds for getting a test question correct at some point during the study as evidenced by their journal writings. However, most could not articulate how to actually revisit the text. Interestingly, we discovered that the strategy of reading the sentence above and below specific information in the text was critical to understanding test questions that required analysis at the sentence level.
MARCH FINDINGS

At the beginning of the study in March, five of the six focus students, with the exception of Jenaya, demonstrated some degree of ability to revisit the text prior to selecting an answer. However, only one of these focus students, Nathaniel, an advanced focus student, actually cited evidence from the text instead of merely from memory. For example, on a question that asked, “What makes the preceding sentence ironic?” Nathaniel said, “Looking back at the paragraph, it says that Elizabeth had been climbing for a long time. But um, she can’t even make it with her friends.” Nathaniel identified key parts of the text and then considered whether or not the answer was correct based on the evidence that he found. He often identified specific parts of the text before deciding on his answer.

The remaining four focus students tended to rely on memory and, therefore, they weren’t as specific in their answers. For example, Rashawndra, an advanced student, on a question that asked, “What is the purpose of the article?” said, “I think the answer is ‘D’ because like when I read the article it’s very informational. Certainly not that interesting but there’s a lot—like they explain facts and stuff about corporations and government.” Although she cited evidence from the article, this was merely from her memory of just reading. She didn’t necessarily go back to the text and double-check the way that Nathaniel did. Reynaldo, a basic test-taker, also tended to respond to a vegetarian question from memory, stating, “... the article says to eat both vegetables and meat although I prefer a lot of meat.” Alejandro and Bill did the same.

Jenaya, a far below basic test-taker, on the other hand, did not mention the text in any of her five responses. Rather, she made statements like, “I chose A because it best to summarize the information in the article.” However, she did not mention any specific text
from any of the reading excerpts. Early in the study, it became apparent to me that most focus
students, with the exception of Nathaniel, were not actually going back to the text prior to
arriving at a decision. Rather, they were relying strictly on their memories. Therefore, during
the discussions, I required students to double-check their responses by going back to the text
and specifically citing evidence during the classroom discussion sessions.

During these discussions, I stressed the idea of revisiting the text prior to making a
decision, particularly when students were trying to choose between two remaining answers.
As I did this, students collectively struggled, at times, to find the necessary evidence in the
text as the following conversation illustrates from an advanced class concerning the test
question from "The Incredible Northern Lights" that asked, "Which of the following details is
supported by details or evidence in the essay?"

Teacher: Okay, so A and D. Some people are not aware that there is a
scientific explanation for the northern lights, or D, the northern
lights can only be seen when the temperature is at its coldest. So
what can we do now?
Male Student: Keep reading.
Teacher: Okay.
Female Student: You go to the answers and you try to find evidence for each of
them.
Teacher: Okay, let's do that. So you said B and C were knocked out.
What about A and D?
Male Student: A isn't really supported.
Teacher: What is it that is not supported?
Male Student: I think it is supported—like, I think it is because after the second
paragraph, it says, "The northern lights are better understood by
scientists." So that could mean that these people are wrong and
there's actually a scientific explanation.

The male student in the above example was able to cite a very specific bit of text, the
sentence that says, "The northern lights are better understood by scientists." He then inferred
that the general population could be wrong about what they thought and that there is in fact, a
scientific explanation that holds greater weight. After considering this option, I decided to first have the students explain why it could not be other alternative before we decided on A.

Teacher: ... Real quick about D [The northern lights can only be seen when the temperature is at its coldest], why isn’t it right?
Male Student: Because the text doesn’t support it.
Male Student: ... A is like more relevant to the article.
Male Student: A is more relevant to the entire article.

In order to be sure that A was correct, we had to be absolutely certain that D could not be correct, and since there was essentially no mention of it in the article, it was safe to conclude that the answer was A. As the remaining male students pointed out, A was more relevant to not only the article, but to the entire article. Students were starting to recognize that in order to pick the most relevant answer, they would most likely have to revisit the text instead of simply relying on memory. This idea of revisiting the text instead of merely relying on memory became apparent in their journal responses too.

One basic student stated, “It was easy to mess up because each answer was a piece from the text but you would have to read and reread and go back and check, make sure.” Here, this student discovered that some answers did involve the text but that they weren’t necessarily the correct answers, and to get an answer correct, one would have to reread the text, to “check” and “make sure.” A below basic student stated, “The next person who reads question 11 and does not get it, then they should reread the text until they start to pick something up.” This student realized that it might take more than one attempt at revisiting the text until the information sunk in, until, in his words, he “pick[s] something up.” Instead of reading being an activity in which students read and were finished, it was starting to dawn on students that they might need to reencounter the text several times in order to arrive at the correct answer. As an advanced student noted, “If someone was to see this question again...
They could think which answer is the most supported and not be confused by multiple right
answers.” Here, the student understood that his confusion could actually be minimized by
revisiting the text and selecting the answer that had the most evidence, or was in fact, most
correct. Students recognized that they should confront the text over and over again in order to
double-check or make sure that they had, in fact, selected the correct answer.

APRIL FINDINGS

During the next phase of testing in April, the focus students made gains in revisiting
the text before deciding on an answer. However, these gains varied among students. One
student, Nathaniel, an advanced focus student, revisited the text consistently before making a
decision and he cited specific evidence. For example, on a question that asked, “Which
statement BEST describes what happened in the story?” he said:

It’s not D [Two girls are not sure how they will cross a creek until a group of
teenagers helps them find a stairway]. That’s not a story . . . . It’s not B [Ashley’s
mother warns her to come home before the high tide]. B is missing the whole tide
thing. The tide was major part of the story and it explains why—the mood—how
they find supplies and why they have to leave. C [Two girls cross a shallow
stairway] is also missing the tide so it has to be A [Two girls get into trouble when
the tide comes in while they are asleep on the beach].

Here, Nathaniel understood that the tide was a major portion of the story, that it was the
problematic event that got the girls into trouble. He equated the tide with mood and
determined that it was a critical part of the story.

Unlike Nathaniel, most of the other focus students cited general evidence from the
text. For example, Rashawdra, an advanced test-taker, responded to a test question about
“Springtime in Washington” the following way. She said:

I think the answer is hmm. Okay, wait a minute; I have to look back at the story
thing. Oh, so yeah so I think it’s C it is impossible to predict exactly when
spring’s warm weather will arrive because that’s what it says in the story. And about the other answers, this one was specifically about the question like when the festival was held I guess. And the other ones were just kind of stuff they didn’t say in the story.

Here, Rashawndra determined her answer because of “what it says in the story,” but she didn’t identify what specifically the text said. And when she ruled answers out, she did so because they were “stuff that wasn’t said in the story,” yet she didn’t specifically identify what this “stuff” was.

Other focus students showed this same lack of specific referencing. For example, on one question, Bill, a proficient test-taker, said:

I’m choosing A because answer B only refers to the first part of the story. Answer C does not refer to another part of the story, and answer D only refers to another part of the story but not the entire story so answer A is the best choice.

While Bill seemed to recognize that the correct answer was the answer that encompassed the entire story, he did not specifically state why the other answer choices only revealed a portion of the story, or no part of the story at all. He was non-specific. Other focus students, such as Reynaldo and Alejandro, showed this same pattern.

Finally, Jenaya did not seem to understand the test questions well enough to even cite specific evidence. For example, on the question from “The Iowa Coast” that asked what the author meant by adopted home, she said, “B Iowa, C Long Beach, D Southern California is not—that’s referring to the adopted home, adopted home for Africa.” Here, Jenaya seemed confused by the term adopted home equating the following place with Africa, the country that she immigrated from. She determined that the answer was Bixby Park but this was the place where the former Iowans met up and did not live. Again, the language barrier for Jenaya seemed to be a severe roadblock.
The focus students’ ability to identify specific evidence in the text was lacking for the most part. With the exception of Nathaniel, the other focus students tended to cite general evidence that came from simply reading the passage or possibly scanning the passage before making a decision. Most of the focus students did not seem to be able to really zero in on specific evidence from the textual excerpts to inform their decision-making process. Therefore, during the discussions, I recognized that I needed to help students to navigate the text, especially on very complex questions.

During the discussions from the second test, students sometimes struggled with how to approach the text in order to find the answer. On questions involving a summary of the whole article, they needed to access to the text quickly yet thoroughly in order to make a good decision as the following conversation indicates from a basic class on a question from the article entitled, “The Polar Bears of Churchill”:

Male Student: Which of the following best summarizes the information in the article?
Teacher: So let’s read the choices.
Female Student: A, every October, polar bears visit the town of Churchill in the Artic Circle; B, polar bears can get sick if they eat from garbage cans; C, when polar bears visit the town of Churchill, the residents should know how to treat them; D, although polar bears are cute, they can be very dangerous if not handled properly.
Teacher: Very good. So again, what’s the first thing you would do here?
Female Student: Look back.
Teacher: If you look back, it’s pretty long. So say you have testing fatigue and you can’t handle reading every bit . . .
Female Student: Scan it.

Here, the student identified the strategy of scanning as a way to get reacquainted without having to reread every word. At this time, we read the first sentence of each paragraph and then the following conversation continued:
Teacher: ... So based on that quick recap, what can you get rid of?
Anything?
Female Student: B and D.
Teacher: Start with B. Why B [Polar bears can get sick from garbage cans]?
Female Student: Because it doesn’t say anything about Churchill . . . .
Teacher: Yeah, it’s kind of like whoa, out there, and it doesn’t say in the town of Churchill . . . . What else would you get rid of?
Female Student: D [Although polar bears are cute, they can be very dangerous if not handled properly].
Teacher: Why?
Female Student: Same reason.
Teacher: ... It doesn’t mention Churchill. Good.

Here, the student rationalized that because neither B nor D mentioned the town of Churchill, these answers were most likely not correct. The question asked, “Which of the following best summarizes information in the article?” and clearly the answer should involve the town of Churchill and not be just a statement about polar bears in general. Once we got it down to two, the following conversation continued:

Teacher: A or C? Which would you get rid of?
Female Student: C.
Teacher: “When polar bears visit a town in Churchill, the residents should know how to treat them.” Why would you get rid of that?
Female Student: It doesn’t summarize it.
Teacher: How confident are you that it’s A?
Female Student: Pretty confident.
Teacher: The answer’s A. So your idea that the incorrect answer probably wouldn’t have the word Churchill in it worked here. I wouldn’t say that you’d want to do that all of the time because there may be a main statement in here that characterizes the whole article without using the town. But that was a good hunch and it worked for you. Do you agree that, “Every October, polar bears visit the town of Churchill in the Article Circle?” Does that sound like a summary of the whole thing?
Female Student: Yes.

Here, the students were able to select an answer that was a general summary of the article.

However, they needed to scan the text in order to be sure that they could pick out a general
summary statement. Two of these statements were too general and did not include
information about the town of Churchill. They were picking up on the idea that a summary of
an article would be general but not too general.

This idea of scanning the text in order to get the general idea became apparent in their
journals. For example, a below basic student stated:

If someone missed the test question, I think it’s cause they didn’t read the question
good or go back and scan the text . . . . I think if someone see it again, they should
read the question and the text good and then scan the text.

Here, this student pointed out that it’s important to return to the text and scan it before
making a decision. A basic student put it in a slightly different way. She said, “I will get it
right by going back and checking the paragraphs.” Here, this student equated double-
checking the paragraphs within the text as a way to insulate herself from choosing the wrong
answer. Finally, an advanced student stated, “Someone who got it wrong could reapproach
the question by looking for text in the story that best disproves their choice.” Here, this
advanced student equated not only reviewing the text as a way to get the correct answer, but
to actually use the text to refute the other answer choices. This advice is very useful for a
question that asked students to select the answer that best summarizes the passage because
students can often be fooled by statements that are too general as the above discussion
indicated. Thus, through the classroom discussions, students were seeing the value in
revisiting the text although while some students were only beginning to see the importance of
scanning before making a decision, others were much more sophisticated and saw the text as
a goldmine of information, a source that they could capitalize on in order to disprove
incorrect answers.
MAY FINDINGS

One focus student showed a pattern of using specific evidence from the text to make his decisions, three showed a pattern of using general textual evidence, and two showed a pattern of simply relying on memory, never really delving back into the text. On the last test, Nathaniel was the only focus student who methodically disproved all other answers before making a decision, and he did this by using textual evidence consistently. For example, on a question from “The Polar Bears of Churchill” that asked, “Which of the following BEST summarizes the information in the article?” he said:

Um, I would not choose B [Polar bears can get sick if they eat from garbage cans] and D [Although polar bears are cute, they can be very dangerous if not handled properly]. D is only answered in one paragraph, the third paragraph. And um, B is mentioned once. Well, C [When polar bears visit the town of Churchill, residents should know how to treat them] is not the main point of the article. And same as D. But A [Every October, polar bears visit the town of Churchill in the Arctic Circle] is the best because it describes the setting, the time in October and the setting, which is part of the main idea, Churchill. And the whole article is about the polar bears and their relationship with Churchill . . . .

Here, Nathaniel seemed to understand, conceptually, that some of the answer choices did not fit because they were only mentioned in a cursory way and therefore, he determined that they did not best summarize the article. He felt that, of the remaining two choices, one was not “the main point of the article” and the other answer was the best because it described the setting, a setting that encapsulated both the time of October and the place of Churchill. He also added that “the whole article [was] about the polar bears and their relationship with Churchill.” Clearly, he not only gave ample evidence of why the answer should be A, he also refuted other choices before making his decision and he established this pattern during all of his May focus test questions. Nathaniel did not leave any stone unturned, and he was the only focus student to demonstrate such a high level of mastery.
Other focus students illustrated this process of revisiting the text by citing specific evidence but to a lesser degree. For example, Bill, a proficient test-taker, seemed to make great gains in citing specific evidence but he didn't rule out all other remaining answer choices by checking back in the text. Rather, he made many statements like, “Looking back at the article, I find, I see that this author of the second article wants you to find out what your individual dietary needs are so I am going with B,” or on another item, he said, “Looking back on the text, I see that C is the right answer because it says in the text that the cherry blossoms bloom a different day each year . . . .” Bill became much more specific before making decisions than he did previously and now he was really extracting textual evidence before deciding on an answer. But still, like Reynaldo and Rashawndra, he was not exhausting all of the answer possibilities.

Finally, the remaining focus students, Alejandro and Jenaya, both below basic test-takers, struggled to cite specific text. Alejandro’s transcripts were filled with statements like “I think it’s D, the northern lights can be seen when the temperature is at its coldest. I think it’s that because it seems most like, it seems more detailed. And it seems like the best answer.” On another item, he stated, “I think it’s C because it gives more evidence to support the answer.” However, he did not explain how C gave more evidence. Clearly, Alejandro made few visits back to the text. He responded primarily from memory.

Jenaya showed this same pattern, but even to a lesser degree. She simply read the question and the answers and then gave a rationale that was directly from the test question. For example, on a question that asked, “Which is the MOST effective revision of sentence 1?” she said, “I chose answer C [Be sure to organize your short story by inserting it into a colorful folder] because ‘a good short story with a well-developed plot must be thought of by
you first.‘’ This rationale was actually from answer choice, A, which read, “A good short story with a well-developed plot must be thought of by you first.” Jenaya was literally borrowing a rationale from other answer choices to back up her decisions. This was a puzzling way to respond to the test question but it revealed that she probably really did not understand what she was reading within the actual test question and therefore, she was unlikely to return to the text to try to establish more clarity.

The focus students in this study showed very interesting patterns with regard to revisiting the text. One gave a very clear rationale and revisited the text consistently. Not surprisingly, this student got all of the focus test questions correct on all three tests. Clearly, his method of exhausting all answer possibilities by disproving them worked. Others like Rashwandra, Reynaldo, and Bill showed mixed patterns. They sporadically revisited the text but mostly it was only to confirm their answer choices and not to rule out other answer possibilities. Bill showed the most growth in this way because early in the study, he did not cite specific evidence at all. He was now citing evidence to defend his answer choices but not to omit other answer possibilities. Finally, Alejandro and Jenaya struggled to revisit the text. Alejandro seemed to think he understood that some answers were more detailed or had more evidence but he usually did not check back with the text. Jenaya, on the other hand, seemed to struggle so much with basic comprehension that she often used other answer possibilities to defend the answer that she chose, a process that indicated that she probably was not understanding what she was reading.

Because the focus students were all over the map when it came to revisiting the text, it became clear to me that during the classroom discussions, I needed to show students more ways to revisit the text. As several of the focus students illustrated, going back to the text to
seek a rationale was very hard work and even though many of the focus students were motivated, they seemed to lack this skill. Therefore, it seemed to make sense to show the students other ways of revisiting the text besides scanning. While this approach worked well for questions that asked students to summarize an article or to explain the author’s tone or perspective, this approach would not suffice for complex questions. So I began to introduce students to the idea of reading above and below a sentence in question in order to extract the actual context. An example of this approach can be seen from an advanced class based on the reading entitled, “The Pros and Cons of a Vegetarian Diet”:

Teacher: So, what’s this question?
Male Student: Read the following sentence from the passage. One of the most powerful draws of switching to a vegetarian diet is the idea that because of the lower cholesterol, vegetarians run a lower risk of cardiovascular disease. This is much like the idea that tying a dirty sock around your neck will cure a sore throat.
Teacher: And what’s the question?
Male Student: What is the author trying to accomplish in this comparison?

After we determined that this sentence from the passage could be found in the third paragraph, we decided to read above and below the sentence in question in order to establish more context:

Teacher: Find it [the sentence in the text]. Where is it?
Male Student: Third paragraph.
Teacher: All right. You guys remember when you’re trying to pick an answer and they give you a little passage, what’s a good rule of thumb when you’re looking for context?
Male Student: Read up and down.
Teacher: So if you read up a sentence and down a sentence, then you make sure that you have the right context. Well, here they are, “Vegetarians must seek other means of receiving these vital nutrients, such as vitamin supplements in order to ensure they receive all they need.” And then, here’s the passage, “One of the most powerful draws to switching to a vegetarian diet is the idea that, because of the lower cholesterol, vegetarians run a lower risk of cardiovascular disease . . .” Then the author is saying,
"This is much like the idea that tying a dirty sock around your throat will cure a sore throat. In fact, the body needs cholesterol and there’s no proof that the dietary cholesterol found in meat has any effect on cholesterol levels associated with heart disease. The secret to maintaining a healthy diet has little to do with whether one pursues a vegetarian lifestyle or an omnivorous one."... So the question says what?

**Male Student:** What is the author trying to accomplish?

**Teacher:** Yeah, you have to forget about what you think and ask yourself, “What is the author trying to do?” This is hard because you have to separate from being the author.

On this particular test question, students really had to revisit the sentences above and below the test question in order to get a better idea of what the author was saying. However, at the end of this discussion, after we had narrowed the answers down to two, it became apparent that the correct answer was that the author was actually “mocking” the idea that vegetarians have lower cholesterol and that students couldn’t really know that unless they read up and down because the text confirmed that the body needs cholesterol and there’s no proof that dietary cholesterol found in meat has any effect on serum cholesterol levels. Yet even though we arrived at the decision that the correct answer was that the author was “mocking,” an idea instead of making it seem “slightly foolish,” a student asked the question, “Couldn’t mocking be too strong of a word to use?” It was a very good question and based on the fact that only one student in the class got the test question correct, it led me to make the following response:

**Teacher:** You know it’s not a great test question. And obviously, if only one person got it right, that means right there that it’s not a great test question. So here you are, trying to pick between B and D and you’re like, “I don’t know if he’s mocking. He’s trying to make the idea seem foolish, so should it be slightly foolish or mocking?” Like neither seems to completely fit, hence, this is why most of us missed it.
This conversation led me to think that no matter how many strategies I gave students; sometimes it could be very difficult for them to identify the correct answer, particularly if the shades of meanings between words were extremely close.

On a subsequent test item related to the same article that we had just discussed, students still struggled to remain vigilant about what the author was saying; it was very difficult for them to resist superimposing their own perspective onto the text. This seemed to be an interfering factor that would not go away as the following conversation confirmed:

Teacher: . . . Based on the second article, which of the following statements is true?
Male Student: I would say A [A vegetarian diet is the best way to a healthy lifestyle] because a vegetarian diet is the only way to go.
Teacher: And what’s the author’s tone? Is he saying that a vegetarian diet is the best way to a healthy lifestyle in the con [second] article?
Female Student: It is.
Teacher: Yeah, but you’re thinking for you . . . based on the second article and not on you. That’s what tricky about these tests is that too much thinking can undermine you. Now, there’s sometimes when your own thinking is good. But, in a situation like this, they’re not asking what you think, they’re asking about which article?
Male Student: The second one.
Teacher: The con one. So in the con article that we just answered the other question from, is he saying that a vegetarian diet is the best way to a healthy lifestyle, is that his tone?
Male Student: No.
Teacher: Not at all, good. So you’re right, that’s way off. B, C, or D, what can we get rid of?
Male Student: B.
Teacher: An omnivorous diet is the best way for everyone to be healthy. Why would you get rid of that?
Male Student: Because he’s saying it’s the only way to go.
Teacher: The only way to go for whom?
Male Student: Like for everyone.
Teacher: It’s a very, very general statement . . . this is the extreme. Do you see how B is kind of extreme?
Male Student: But isn’t that with both foods, not just meat—it’s meat and vegetables.
Teacher: Yeah, I don’t know if he’s saying that everybody should eat meat and vegetables because if you’re saying that everybody should eat meat and vegetables, that’s extreme. Does he say “everyone” in there? Whenever you see an extreme statement like “always” or “never” or “everyone,” it’s typically not going to be correct because it’s too—it locks you in too much.

Even though the author did not seem in favor of a vegetarian diet, he did not state that everyone should eat meat and vegetables. However, when students saw this statement, many of them felt that it was the best answer, the truest statement. It was difficult to convince them otherwise so I had to explain that typically, when extreme statements are used on standardized tests, they are not the correct answers:

Teacher: So we’re down to C and D. What do you think?
Male Student: Get rid of D [The perfect diet has yet to be discovered].
Teacher: Why?
Male Student: Because it may say in the article that the perfect diet has yet to be discovered, but it also says in it facts that liver and kidney chemistry should determine what a person should eat.
Teacher: What’s the question asking?
Male Student: Based on the article, which of the following statements is true?
Teacher: So, it’s not asking you what you think is true; it’s asking you what the author of the second article says is true. There’s a separation—a difference there. If you look back at it, what does it say about liver and kidney chemistry? . . . He’s not saying that everybody should do one thing, “Individuals should plan vegetarian diets according to their specific needs.” A very, very tough article . . . . The actual answer is C [Liver and kidney chemistry should determine what a person should eat].

It was clear to me, at this time, that many of the students had a difficult time comprehending this article. It contained many random facts and the nature of the article required students not only to comprehend the different sentences, but to infer and piece together what the author meant. The context of the article was obscure at times and not like other reading excerpts that we had read which seemed to be much more straightforward. Therefore, I added the
following commentary to help them understand where they might have gone wrong and to reassure them that it was, in fact, a difficult text:

Teacher: And if you look at what you picked . . . mostly B actually. So somehow you thought, in your own mind, maybe that if you picked B, you were probably thinking on your own that a diet with meat and vegetables is best for everyone, but you’ve got to watch that word “everyone” and you’ve got to also watch that you’ve got to go with his perception and not yours. It’s tough.

Finally, my closing words led me to something I thought I’d never say to my students in an educational setting, but I wanted to help their spirits; I wanted them to know that even if they superimposed their own ideas, it was not a bad thing. It was just a dangerous thing to do on a standardized test.

Teacher: It’s almost like I’m telling you not to think for yourself when you’re taking this test and that’s like such a horrible thing to tell you. But, at the same time, you’ve got to understand that these tests are not written for you individually, they are written for the masses—they’re standardized. So when you go after the questions, you’ve got to be very, very careful to look at every single statement and to try to figure out what they’re asking you.

The reading examples provided from May were taken from the advanced class just to illustrate how complex the process of revisiting the text could be for all students involved in the standardized testing process. And it was clear to me that if my advanced class had this much difficulty, the degree of difficulty for my other classes was even higher. And the journal entries confirmed this finding. Below average students tended to make comments that were very general when it came to reading the text. For example, one student said, “I could reread the text and read it better and make sure I read the words right and the text.” However, I wasn’t sure, at this point, how I could truly redirect a student who was struggling with reading to master the types of test questions that my advanced students were falling down on.
Basic students echoed the vagueness that many below basic students demonstrated when writing about revisiting the text. However, sometimes, basic students were able to couple revisiting the text with other strategies. For example, one basic student wrote, “If I were to see this question again, I would probably reread over the two answers I got it down to and then reread the article as well as the question. Then I would choose the best answer.” While this student explained that he would simplify the ordeal of revisiting the text by first narrowing down his choices to two, he didn’t mention, in any detail, how he would reread the text. For example, would he read up and down if it was a question that included a sample sentence? Would he skim and scan if it was a question that asked him to summarize? Like the below basic students, when mentioning revisiting the text, the basic responses were vague.

Advanced students, however, seemed to internalize how to revisit the text because they were mentioning specific strategies in their responses. For example, one advanced student stated:

I missed this question because I didn’t choose the most correct answer and I chose my opinion over the author’s. If I saw this on another test, I would eliminate the incorrect answers and try my best to fully feel the author’s tone.

Although this student gave a better description than the others regarding how he made errors because he chose his opinion over the author’s, he didn’t really explain how he would revisit the text. Rather, he stated that he’d try his best “to fully feel the author’s tone.” In conclusion, when it came to revisiting the text, most students seemed to have a difficult time explaining precisely how they would do just that.

One of the 37 students, a basic focus student, did not mention revisiting the text at any point during his journal entries. However, this student left several entries blank as he
struggled to articulate his ideas in writing. Interestingly, he showed signs of revisiting the text in many of his think aloud entries, only he was sporadic in his approach; sometimes he’d explain specifically why he eliminated certain answers by using specific textual evidence, but most times he would not.

**STUDENT SUMMARY RELATIVE TO REVISITING THE TEXT**

The students in this study showed strong signs of learning to value revisiting the text but the extent to which they could explain this left something to be desired. Although throughout classroom discussions, students provided textual rationales as to why they would get rid of certain answers, this practice did not appear to resonate with them completely as evidenced by many vague think aloud transcripts and journal responses. Revisiting the text is difficult work that takes many forms. As seen through the classroom discussions, sometimes it requires scanning the text to look for general ideas while other times it requires specific adherence to particular sentences or paragraphs within the text. It was challenging for most students to know how, precisely, to revisit the text. The greatest challenge that students seemed to face with regard to revisiting the text was limiting their own knowledge, and not allowing it to outweigh what the textual evidence actually said.

**Integrating the Strategies General Results**

Results of the study indicated that students’ ability to read and interpret the test question, eliminate answer choices, and revisit the text concurrently appeared to be articulated by about one-third of the students. By the end of the study, only one focus student, Nathaniel, showed strong signs of how to blend all of the strategies while he tested. Through
the discussions, 11 out of the 37 students in this study wrote that integrating all of the strategies together was the best way to increase their odds of getting the test question correct as evidenced by their journal writings. Interestingly, we discovered that although integrating the strategies concurrently was complex and difficult work, the degree of integration often depended on the test question at hand. In addition, an individual’s reading ability was often a prerequisite skill, but not a guarantee, necessary for mastery.

**Focus Students’ Abilities to Integrate the Strategies**

Throughout the study, all focus students improved at reading and interpreting the test questions, eliminating answers, and/or revisiting the text; however, the degree and extent that they were able to practice these skills concurrently differed considerably. For example, Nathaniel, an advanced student, was able to practice these skills almost effortlessly. By the end of the study, he knew how to look back into the text and he knew how to zero in on what the text question was asking. For example, on a question asking, “Instead of following a strict vegetarian diet, the author of the second article encourages students to ____________,” Nathaniel said, “To answer this, I will go back to the article into the second argument.” He paid attention to the question and noticed it asked about the second article and not the first. Then he stated, “And um, I would look in the last paragraph in the conclusion where the author would sum up . . . .” Because the question asked about what the author encouraged, Nathaniel thought that this information would be summed up in the last paragraph. He had an automatic sense of where to find either confirmation or non-confirmation of the choices. However, when the last paragraph didn’t tell him everything he needed to know, Nathaniel refined his approach, continuing to actively engage while he searched. He added:
So if your position is to find out what the perfect diet is... there’s the fourth sentence in the second paragraph, “Everyone should be medically assessed by a medical professional in order to find out what is the best diet for him or her.”

Nathaniel knew he needed to know what the author encouraged. He appeared to think he’d found it in the line, “Everyone should be medically assessed by a professional...” because he concluded, “So knowing the supporting text, I would lean towards B [find out what his or her individual dietary needs are].” Then he went onto explain how the other choices didn’t really back up what the author was saying. Compared to the other focus students, Nathaniel had a very strong sense of where to go in the text and that influenced his ability to practice the other strategies of eliminating answers and reading and interpreting the test question.

During post study interviews, when asked, “Tell me how you approached a typical test question,” Nathaniel had the following to say about his test-taking processes:

... I look at the question first... and then I go back and I look at the answers and I get a feel for the one I think is right... Then I look back [at the text] and I try to support it or find ones to eliminate... I usually look for the one that has the most information about it. Like, the one that has the most written about it, like, the one that the author focuses on most....

Here, Nathaniel mentioned the importance of really looking carefully at the test question and understanding what it was asking before proceeding. He then expressed that although he read the answer choices to get a feel for the right one, he then followed up by looking back in the text to see what answers he could find support for and which ones he could get rid of. He tended to choose those that had the most information, or the choices that the author focused on most. He also tended to have a sense of the overall reading excerpt before selecting an answer. Finally, he added that just looking at the answer choices and making a determination based on that alone could prove to be hazardous. He added:
Just looking at the answers gives you a very vague picture of what you’re looking for, and I think you’ll get kind of thrown off by not looking at the question first. If you’re not meticulous in your process, you can miss it [the test question] on the first look . . .

Here, Nathanial realized that reading the test question carefully is immensely important or you can miss it on the first read. He also knew, based on the previous excerpt, that when eliminating answers, it's especially crucial to follow up in the text and to look for ways to support or denounce answer choices. He then added that if you don’t follow such a process meticulously, you could easily miss the test question. Nathaniel, unlike the other focus students, seemed to have a real system of how to approach the test questions.

Other focus students made less clear comments when being interviewed about their processes. For example, Rashawndra said, “I read the question and then I read all of the answer choices, and then I went back to the text.” Rashawndra didn’t talk about what happened after she narrowed it down to two. She also never mentioned revisiting the test question again. Bill said, “It was pretty simple to read what was going on . . . I just put the question in my mind and then I found the answer that best-fitted.” Bill seemed to underestimate the complexity of some of the test question characterizing the experience as “pretty simple.” Finally, Reynaldo said, “I read the question first and then I look at the choices, and then I look at whether I need to read a paragraph or just scan the whole passage.” Reynaldo commented on taking a minimalist approach; he said that he didn’t like the experience of testing so he chose to do the minimum to he could to get to the answer.

Interestingly, neither Rashawndra, Bill, nor Reynaldo commented on how they’d go back to the text. They, unlike Nathaniel, didn’t mention the importance of reading the test question extremely well, or the importance of not getting too distracted by the answer choices
for fear they'd get thrown off. Unlike Nathaniel, they didn't seem to have a sense of how to find the answer that the author gave the most information on, or the answer that the author focused most on. These intermediate focus students remained somewhat vague in their descriptions and this may be partially why they did not crossover into the realm of using the strategies concurrently. Their descriptions were all linear stating they'd do one thing, like reading the test question, but then never mentioning it again.

Alejandro, a below average test-taker, and Jenaya, a far below average test-taker, described their processes even less clearly. Alejandro said, "It wasn't really hard ... I just read the test question and then I read the text ... I went through answers and I crossed every which one out that didn't make sense." While Alejandro had a system of reading the question, the text, and eliminating answers, he also perceived that it wasn't very hard yet he wasn't very accurate on his tests getting less than half correct on each one (see Table 8). In addition, Alejandro explained that he read the test question before the readings and then he never mentioned reading the test question again. Perhaps, he did not clearly remember what the test questions asked by the time he finished the readings and this may have interfered with his ability to get items correct.

Jenaya, on the other hand, stated:

I read the question first. Then I read all these [the answer choices] and then I went back to the text. I started reading and looking for the answers ... I crossed out which answers are not there and I didn't cross out the one that was there.

Although Jenaya seemed to have an idea of how to approach the test questions, she struggled with understanding the actual text. Regarding a passage on bamboo, she said, "It had like long paragraphs and it's hard to read." In addition, Jenaya did not mention reading the text first; rather, she described reading the question and then the answers and then looking in the
text to see what she could cross out. This approach would most likely prove ineffective for
even the strongest of readers because eliminating answers without fully comprehending the
text would most likely lead to the elimination of correct responses. Hence, even though
Jenaya did learn a process of what to do, her strategic approach was flawed. Because she did
not read through the readings initially, she was likely to misunderstand since she was
primarily looking for the answer and not seeking to comprehend the meaning of the passage.
In addition, because she already had many reading comprehension issues, her ability to be
accurate was greatly compromised.

Overall, the focus students improved individually over time on various reading
strategies, but when knitting them together in a sophisticated manner, only Nathaniel seemed
to prevail. Some reasons that the others did not was because they may have been taking a
linear approach vs. a cyclical approach, practicing a strategy once and then abandoning it.
Perhaps, overconfidence could have also interfered as both Bill and Alejandro mentioned
how easy the process was. In addition, motivation may have also played a part as in the case
of Reynaldo who did not appear motivated and chose to take a minimalist approach. Finally,
a student’s reading comprehension level appeared critical. Jenaya already struggled with the
difficulty of the texts so adding this multiplicity of strategies may have simply overwhelmed
her. There was a stark contrast between Jenaya, who seemed to constantly misinterpret the
text, and Nathaniel, who seemed to automatically see the big ideas that the author focused
mostly on.
JOURNAL REFLECTIONS CONCERNING INTEGRATING THE STRATEGIES

During the journal reflections, 11 of the 37 students mentioned the importance of integrating all of the strategies prior to making a decision on a test question. Interestingly, the students who mentioned integrating all of the strategies simultaneously came from all of the classes. For example, an advanced student said:

... The question specifically says “evidence in the essay” but someone could have definitely not read the question thoroughly. How someone can now approach the question differently is they could read the essay ... look at the text clearly ... or read the question and all of the answers all the way so they know exactly what their task is and choose the answer that is backed up with evidence. Also, if stuck on two choices, pick the one that is provided with more evidence than the other.

Here, this student really seemed to see that she needed to really understand the nature of the test question before proceeding to the next step. She mentioned that the question “specifically” says evidence in the essay, and that someone could have misread it if they did not read it “thoroughly.” She described reading the text as “look[ing] at the text clearly” and then reading the answer choices “all the way” before deciding. Finally, she identified the most desirable answer as that which had the most evidence from the text. This student seemed to understand that all of the strategies were integrated. However, out of her 11 classmates, she and only two others mentioned the integration of all of the strategies. Even Nathaniel, a star focus student, only mentioned a maximum of two strategies in his journal responses. However, this may have been because during some of the discussions, we only focused on one or two strategies. For example, on questions that asked about summarizing, we tended to focus on how to eliminate answers and how to scan the text. We did not focus on how to read or interpret the test question because the nature of the question, “Which sentence best summarized the article?” was straightforward and seemed obvious.
While some basic students, 6 of 17, mentioned integrating all strategies, their responses were also quite insightful, as they knew that they were experiencing a cyclical process, not necessarily one that was strictly linear. For example, one average student said, "If I were to see the question again, I would probably read over the two answers I got it down to and reread the article as well as the question. Then, I would choose the best answer." Here, this student recognized that after narrowing down the choices to two, he would reread both the question and the article again before selecting an answer. Although he was not specific about how he would revisit the text, he was very clear that he would repeat some of the strategies, reading the test question and reading the text, before making a decision.

Finally, in the below basic class, two of the nine students were able to mention all of the strategies. For example, one said, "I got it right because I read the question over and over. Then, I went to the text and scan the text over and over. So I went back to the question and answers and I got it right!" Here, this student mentioned the importance of going over the question and the text multiple times, and that upon returning to the answer choices; he was able to get the test question correct. Although none of the basic or below basic students who mentioned all of the strategies were explicitly clear in their responses (possibly because the journal prompt didn't call for such detail), all of the students who did mention all of the strategies seemed to view the strategies as something to be done over and over in a cyclical fashion and not simply in a linear way.

**STUDENT SUMMARY RELATIVE TO INTEGRATING ALL STRATEGIES**

The students in this study showed mixed signs of integrating all of the strategies concurrently. However, these results should be interpreted with extreme caution. Although
approximately one-third of the students expressed ways to integrate all of the strategies simultaneously as evidenced at some point in their journal writings, this was not necessarily a guarantee that they were able to practice such behaviors simultaneously. Conversely, just because students did not mention integrating all of the strategies, such as Nathaniel, it does not mean they did not practice them while they tested. Rather, the lack of students describing all of the strategies simultaneously may have been because the nature of the discussion on the test questions students wrote about did not demand use of all of the strategies equally. For example, a test question could have been extremely straightforward without any need for interpretation. However, what is more clear concerning the results of students using strategies concurrently is that some students seemed to “see” the test strategies as interconnected, cyclical, and worthy of repetition while others seemed to perceive them as linear, rigid, and to be practiced only once.

**Feelings About Testing General Results**

Results of the study indicated that students’ feelings about test taking changed throughout the study. However, the degree and extent to which their feelings about testing changed was difficult to gauge because most of the students did not mention their feelings about test taking relative to the study. Initially, two of the six focus students, one advanced and one basic, expressed either confusion or frustration while testing. By the end of the study, two focus students, both advanced, expressed more confidence. Many focus students did not comment on their feelings about testing while they tested although end of study interviews revealed that below average focus students had a more difficult time explaining their thought
processes. They also expressed more of a linear process approach when approaching items than that of their higher achieving peers.

Through the discussions, 18 of the 37 students, or approximately one-half, mentioned in their journal writings how they felt either while they were testing or how they felt about the testing intervention in general. Because students were never explicitly asked about this, aside from the focus students during the end of study interviews, it is interesting to note that many of the students in this study chose to express their feelings about testing voluntarily.

**MARCH FINDINGS**

At the beginning of the study in March, two of the six focus students expressed their feelings about testing; however, these expressions took very different forms. Rashawndra, an advanced test taker, expressed both certainty and doubt on specific items, while Reynaldo, a basic test taker, expressed cynicism about testing several times as he worked his way through the five think aloud questions.

Of the five think aloud items, Rashawndra expressed doubt on two and certainty on one. For example, on an item asking, "Which is the most effective revision of sentence 1?" Rashawndra said, "Um, this one’s of tough for me . . . . So, I have to admit, I’m a little confused by this one." Rashawndra expressed her feelings about items and when she wasn’t sure, she expressed that she would have to resort to guessing. For example, on another item, she said, "This one’s hard for me because I have never heard the word integral used my whole life . . . . I think I’m going to have to take a guess on this one." Not only did Rashawndra add commentary when she was confused or unsure, she also made statements when she felt she was accurate. For example, on a question that asked, "Which statement
BEST describes what happens in the story?” she said, “It’s kind of an obvious answer to me.”

On three of the five test questions, Rashawndra expressed her thoughts about her own level of accuracy.

Reynaldo, on the other hand, expressed his feelings about testing, but his commentary had little to do with his accuracy on items. For example, he made the following statement in between two think aloud items:

All right, I’m flipping the ugh pages to the next one. Why do we need to take a test? To move on? It’s pointless. You have the grades, you move on. You don’t have the grades, you shouldn’t move on. That’s the way it should be. Not everyone is good test-takers. Those who do poorly [in school] the majority of them are just great test-takers. Those who do very well [in school] are somewhat poor test-takers. Kind of sounds stupid but it really is the opposite.

Reynaldo’s theory that students with low grades test well and those with high grades test low seemed to preoccupy him while he completed the think aloud questions. He concluded that tests were unnecessary, “pointless” as he put it. He equated value in school with grades and not tests. As he continued through the think aloud questions, he made sporadic comments like, “Oh, I first need to read the test [excerpt]. Let’s just scan it. Don’t take too much time.”

He approached the test questions by doing the minimal amount of work possible in order to get the test question correct. Clearly, he felt that the experience was unpleasant and unproductive.

Thus, early in the study, it became apparent to me that some of the focus students, regardless of their testing level, had very different thoughts about testing. While Rashawndra made comments to gauge her level of accuracy, Reynaldo made comments to express his disdain toward testing in general. Therefore, during the classroom discussions, I stressed the idea that even though test preparation work was difficult and complex and sometimes might
even feel unpleasant, it was worth learning strategies in order to attain more mastery as the following teacher comments illustrates with a basic class:

Almost everybody got the answer wrong so that’s why we’re looking back at these. We’re trying to figure out how we missed it. Did we just not pay attention or did we actually try to answer it?... What I’d like you to learn here is that if you really understand how to decompress the questions and tackle one step at a time, it’s not as insurmountably difficult as it might appear. And you do have the capacity to revisit the text. You do have the capacity to go back to the question. You have that capacity. It’s like these are tools that you need to utilize whenever you’re testing.

Especially early in the intervention, I had to remind students that they were learning skills in order to be successful later on. For many of them, this appeared to feel like bad medicine.

During the early part of the study, I had to model and explicitly teach how to go back to the text to scan, how to interpret the test question, how to narrow down the answer choices. And yet students still had their own ideas about the importance of testing as the following conversation with a below average class illustrates:

| Teacher: | Ima just read question 3. What is the first thing you would do when you see this question? |
| Male Student: | Go back to the story and read—do a scan. |
| Teacher: | Good. Why would you go back to the story and scan? |
| Male Student: | So you could remember it. |
| Male Student: | So you could find the right answer. |
| Teacher: | So you make sure you find the right answer. Now, what is the question specifically asking? |
| Male Student: | The main idea. |
| Teacher: | Now... a lot of times when these tests ask students to get the main idea, the students get confused and get the answer wrong... So if we look at this article and scan it, what’s the main idea? What are they trying to do in this article? |
| Male Student: | I don’t know what they’re trying to do. |
| Teacher: | You’ve got to look. |
| Male Student: | Oh, we have to read? |
Initially, I thought that the students were being sarcastic, but as the conversation continued, I realized that they seemed to have a lot of bad feelings around tests as the remaining conversation illustrates:

Teacher: Well, [you have to] scan. On these multiple-choice items, you have to scan [the text] for what's in there. Is this different than the normal stuff you read?
Female Student: Well, no.
Male Student: I don’t even read. I don’t know how to read.
Male Student: It’s the fact that it’s a test.

As one student pointed out, “it’s the fact that it’s a test” and yet even though these practice tests did not contribute in any negative way towards the students’ grade, students, particularly below basic students, still showed a high degree of discomfort around taking a test in general; one even stated that he didn’t know how to read. Later on in the conversation, I attempted to reassure the students after we talked about the bigger vs. smaller picture among the answer choices from the reading, “Welcome To Our Company/Cake Director”:

Teacher: This one is kind of too small an answer because it’s just talking about the best ways to decorate a cake. So the job description is bigger than that. Do you guys see that?
Male Student: Yeah.
Teacher: If you can start to see through this stuff, you guys can start to nail these tests . . . . It’s so important to take a few minutes and think about how we get caught up on the wrong things when we’re looking on these tests, when really, they are manageable if we just break it down.

Students, at this point, seemed to see some value in persevering in the discussions. However, when it came to the journal responses, the below basic students did not give any details about their feelings about testing, either positive or negative, in their journals and many of them did not complete their entries. Of the nine students, none mentioned that they felt particularly positive or negative with regard to test preparation. Early in the study, however, some of the
advanced and average students, on the other hand, expressed a lack of desire. For example, an advanced student said, “I really didn’t feel like doing the test. I had other stuff on my mind like other schoolwork and I kind of got lazy . . . . I really had my mind set on the STAR testing.” This student did not see the relationship between learning effective test preparation strategies and how those might benefit him on the California STAR test taken in May. He also mentioned the pressure of other schoolwork was occupying his mind. Clearly, he did not see this type of intervention as a priority. A basic student put it another way stating, “I was tired of reading and started to get confused.” At this time, this student didn’t equate how reading more thoroughly and strategically might have reduced his confusion. He didn’t seem to possess a great deal of testing stamina at this stage, nor did he seem to see much value in the intervention.

APRIL FINDINGS

During the next phase of testing in April, the same two focus students commented on their thoughts again only to a lesser degree. Unlike before, this time Rashawndra only commented on one think aloud item and she expressed doubt, indicating she was going to guess. On the item concerning “The Polar Bears of Churchill” that asked, “Which of the following BEST summarizes the information in the article?” she said, “. . . Hmm, I don’t know. This is really hard like. Uh, I don’t know. To be honest, I think I’m just going to guess . . . and say umm, D.” At this point in her learning, Rashawndra saw no other alternative but to guess, even though she had gotten the answers down to two choices.

Reynaldo, at this stage, stopped complaining about having to take a test but he then claimed to “try his best” on an item he got stuck on, yet he did not activate any strategies
whatsoever. For example, on question 10, he said, “Um, I don’t really understand this question. Ugh, I think I’m going to guess. I don’t really get it but I’ll try my best. Well, I’m not really sure. I’ll guess. I think it’s B.” Here, Reynaldo did not even read the test question [Which of the following is supported by details or evidence in the essay?] aloud. He rather instantly assumed he could not answer it. Unlike Rashawndra, Reynaldo did not even attempt to eliminate any choices. However, like Rashawndra, Reynaldo did not seek out additional strategies such as rereading the test question or going back to the text to unearth evidence. While the reasons that these two focus students may have felt unsure about specific items could have been due to them not utilizing many strategies (in Reynaldo’s case, he didn’t use any), the other focus students did not offer any commentary regarding how they felt when they tested so it was unclear, at least at this point, what they actually did at the moments that they weren’t feeling efficacious.

During the classroom discussions, students seemed to be unsure of which strategies to apply and when until, during the discussions, I pointed out that the way in which students approached the test questions would vary and that each student would have to discover his/her own process as the following conversation with a below basic class illustrates:

| Teacher: | So what would be the first thing you’d do after you read through that [Which of the following ideas is supported by details or evidence in the essay?]?
| Male Student: | Eliminate the answers that don’t fit.
| Teacher: | How about you?
| Male Student: | I’d go back to the text and see what it says first and then go back to the question and eliminate the ones that don’t make sense.
| Teacher: | So you’d go back to the text first and then start eliminating.
| Male Student: | Right, scan it.
| Teacher: | So your process is going back to the text and your process is trying to eliminate, so you can see that we don’t all do the same thing. As long as it’s working for you, you should be able to lead your own learning . . .
Early in the study, the below average group seemed to be thinking there was a "correct way" of responding to the initial question, "What would be the first thing you'd do after reading through the test question and answer?" At this time, it occurred to me that students needed to be reassured that their own test-taking process might not mirror that of the person sitting next to them. This idea of students establishing their own processes in the order that they saw fit became apparent in other classes too. In addition, what also became apparent is that getting a test question wrong was not necessarily a bad thing if there wasn't any way to avoid it. For example, a student experiencing a lack of vocabulary suggested a good reason to miss, as this conversation from the basic group exemplifies:

Teacher: ... Omnivorous is really an important word to understand here. What does it mean to you?
Male Student: Everything.
Female Student: That you just eat meat.
Male Student: No, that's carnivore.
Female Student: Oh, never mind.
Teacher: This is a very good point. The beautiful thing about this test prep stuff is we learn more from our mistakes then we do from our correct answers. In school, so much of what we do is try to be correct all the time and we get praise when we're correct yet you're trying to get rid of what's not correct.

As this conversation shows, sometimes students stopped short and did not want to explore why they were incorrect in their thinking. For example, the above female student said, "Oh, never mind" when she was corrected by the male student. However, she recognized that not understanding a word was an important part of establishing a sense of efficacy; it was a good reason to miss a test question as the follow-up teacher comments suggests:

So if you don't quite understand omnivorous, then that's a really good reason of how you could miss the question. That's actually a good reason. If you're going to miss a test question, let it be because you don't get a word. That's a good reason to miss a question because you can't really do anything to fix that. But if you have
some skills and you can apply them to the text, you shouldn’t be missing a lot of the questions. So, very good, you just misinterpreted omnivorous.

Here, the student who misinterpreted the term omnivorous still received some praise because there was little she could do to get the test question correct since she did not understand a key vocabulary word that was part of the test question. I was trying to illustrate the point that it’s okay to miss a test question if it’s from lack of vocabulary. However, what’s not okay is not trying any strategic alternatives on test questions that are clear.

Journal responses showed that a lack of motivation was becoming apparent and this may have also interfered with students’ level of effort to clearly interpret the test question, exhaust all answer possibilities, and seek evidence in the text. Approximately one-third of the journals mentioned some aspect of being too lazy to try on the some of the questions and these comments about laziness came from all classes. For example, an advanced student said, “I started to get lazy and didn’t want to do the test anymore so I just chose any answer.” This student didn’t try as much as she had previously. She seemed to tire of the test-taking experience. Others shared this sense of fatigue as the following comments from a below basic class indicates. One said, “I was kind of lazy and tired and I wanted to get the test done.” Another added, “I misread it [the question] and I was lazy so I got the question wrong.” Finally, a third said, “The day we took the test I was tierd so I geussed a lot.” These students seemed to get distracted by either a sense of laziness or tiredness. There was either a sense of urgency, “I wanted to get the test done,” or lack of motivation to revisit the test question, “I was lazy so I got the question wrong,” or sheer tiredness which led to guessing, “I was tierd so I guessed a lot.” Regardless of what they were learning, many students were not motivated to practice the strategies they were being taught while they tested.
However, a few students did comment on how the strategies appeared to be improving their ability to get answers correct and they seemed to exhibit a sense of pride.

For example, this basic student made the following comments in her journal:

I got the answer right and I think I got the answer right because I eliminated all the wrong answers. I felt that the others just didn’t fit well. I think if I continue to do as I did then I will get it right . . . .

This student seemed to feel a sense of accomplishment from getting the answer correct by eliminating the other choices systematically and she felt that this approach would help her to continue to get answer correct in the future. Another basic student put it like this:

I think someone could have missed this question because they probably might have got it to the 50-50 situation with only two to pick but they might have thought that one sounded better than the other, or seemed like a better answer—and guessed, but did not go back and read what the question specifically asked for. Someone might approach this differently if they see a question like this again. Like me, the first time I didn’t care, I was too lazy to go back and read the text but this time I took the time to double check the question and I got it right.

This student also appeared to feel a sense of accomplishment because he equated rereading the question and the text with helping him to select the correct answer, particularly after he got it down to 50-50. Earlier in the study, he said, “I was too lazy to go back and read the text,” but then after being incorrect, he decided to “double check” the question and the text in order to increase his odds of getting the test item correct. Although rare, these students’ commentaries suggest that for a few, the intervention made them feel more efficacious. Most students in the study, however, did not comment either positively or negatively about the intervention at this time.
MAY FINDINGS

On the last test, two advanced focus students made comments about their testing. This time, Nathaniel expressed confidence on two items, while Rashawndra expressed doubt on an item but then persevered until she felt more confident. This was a change in her behavior. The others, including Reynaldo, did not offer any comments, either positive or negative, during their think aloud testing experience on the last test.

Nathaniel, an advanced student, began to make subtle comments that revealed his confidence on his last test. He made statements like, "Umm, to answer this question, I could plug in the answers, but since I know what this word is, I can choose D." For this test question, Nathaniel was so certain about the meaning of the term "horde," that he did not feel the need to exercise the plug and check method of substitution with all of the choices. Still, in order to be thorough, he eliminated all other possibilities, but he was confident from the start that he knew the answer.

Rashawndra, on the other hand, seemed to feel overwhelmed at times but then she reverted back to strategies that helped to keep her calm. For example, on question 25 that asked, "Which of the following questions could be BEST answered by further research?" she said, "I'm really tired because this test had too much reading in it." Then, she added, "I think all of these could be answered by further research but the key word was best because it's bold and capitalized." Even though she was feeling a sense of fatigue, she noticed the term best and this helped her to narrow her choices. She followed through by going through a systematic process of elimination and she managed to eliminate all answer possibilities except the correct one that she chose. She was learning to fight through her fatigue.
During the last test, none of the other focus students made comments about their feelings about testing. Even Reynaldo, who was so vehemently against testing initially, did not say anything regarding the testing process on the last test. The other students, Bill, Alejandro, and Jenaya, did not say anything about their feelings about testing throughout the entire process, although their end of study interviews suggested their thoughts about testing. Bill said, “I’m a good reader and I have a large vocabulary so it [testing] is pretty simple for me.” Because Bill considered himself to be a good reader with a high vocabulary, he felt confident about the testing process. But he also said he didn’t think it was fair that he had to take so many tests. He stated, “I did a test [The California STAR already] and I didn’t think taking another test [the May test] was fair.” Therefore, even if Bill’s skills were good, his feelings about test taking, and his perception of it not being fair, could have affected his performance. He also added that on the last test, he was “anxious” to be done and therefore he rushed through.

Alejandro, on the other hand, perceived the test readings to be “too difficult and the test readings were too unfamiliar.” He perceived the non-fiction readings like “The Incredible Northern Lights” to be very difficult and other fiction readings like “The Beach Day” to be much more manageable. He said, “… In the other stories, they talked about normal stuff like the plots and stuff like that.” He perceived fiction to be much easier to navigate than non-fiction and attributed much of his difficulty to questions related to the non-fiction excerpts. However, often with fiction excerpts, Alejandro claimed to provide rationale for eliminating answers, but often did not. For example, on a question from the fiction excerpt, “A Visit to the Doctor’s Office,” he made vague comments like, “I think the answer is C because it gives
more evidence to support the answer.” And although he thought he only struggled with non-
fiction, he seemed to struggle just as much with fiction.

Jenaya perceived the readings to be “kind of difficult.” She described the text excerpt about bamboo as having “long paragraphs and it's hard to like read.” She said she tried hardest on the April test but she did not know why. She said that the discussions made her realize there were different ways to approach the test questions, that she could do the strategies in different orders. Yet ironically, her focus student transcripts reveal that she typically read the test question first, then the reading and that she often eliminated answers while she was reading through for the first time. The focus students who seemed to struggle the most with test-taking seemed to inaccurately describe their processes; they often did not actually do what they described.

During the last series of classroom discussions at the end of May, students, for the most part, seemed to tire of testing. They had been through district mandated Practice CAHSEE testing in early March, California State mandated STAR testing in early May, and my intervention tests in mid-March, mid-April, and mid-May. The last series of discussions came at nearly the end of the school year and their patience with testing seemed to be wearing thin. Still, they managed to recall critical information that we had learned inferring that their collective feelings regarding how to test had improved. A basic class discussion from “The Iowa Coast” seemed to sum up the intervention quite nicely:

Teacher: Who can read this one?
Female Student: And for a few moments, we are still trying to dispel the loneliness that so often afflicts them here in their adopted home.
Teacher: Without doing anything more, when you first read that, what are you thinking in your head that you should do?
Female Student: Read the question.
Teacher: Like you’d read the choices, would you do that?
Yeah. Would anybody do anything different?

Yes. You could go back to the text. I would read the choices first and then go back to the text.

Okay, that’s good. Everybody might have a different way of doing it, which is pretty interesting because our minds all work differently. So by “adopted home,” the author is referring to?

At this time, it was important to acknowledge that students might not have the same exact process for approaching a test question depending on their own level of knowledge and the nature of the test question at hand. At this point in the intervention, students seemed to acknowledge their different processes quite naturally; they were no longer trying to convince others of doing it “their” way. The conversation continued:

Teacher: What would you get rid of?
Female Student: I would get rid of—
Female Student: I would have to go back to the text.
Teacher: This is good because this is real. You’re like I would get rid of this and you’re thinking you would go back to the text to make sure.

As we moved further into discussing the test question, once again, the importance of key vocabulary surfaced, as it had all along. The word “former” needed to be defined. And when students came up with differing ideas about the Iowans adopted home, there was seamlessness about it as if that was entirely acceptable:

Teacher: What does this word mean?
Female Student: Former.
Teacher: Former Iowans. What does that mean?
Female Student: They’re not Iowans anymore.
Teacher: And where have they gone to?
Female Student: Southern California.
Teacher: Los Angeles.
Female Student: It’s definitely not what class?
Teacher: Iowa.
Female Student: Because they moved from there. Again, it’s a word thing. If you don’t quite understand “adopted,” then you don’t know what
Adopted home means which home?
Female Student: The one they moved to.
Teacher: So here we go again with vocabulary....

As the conversation continued, it became clear that, paying attention to key vocabulary was crucial, and yet, as I listened to their conversation, there was an ease with which they spoke now, they were less eager to point out the frailties in each others’ responses. The conversation went on:

Teacher: You say it’s not A, Bixby Park, how come?
Female Student: Bixby Park is in Long Beach and that’s where they’re meeting.
Teacher: Okay, do you think every single Iowan moved to Long Beach?
Male Student: No.
Female Student: No.
Teacher: Yeah, so do you guys recall in the very beginning that we talked about big picture vs. smaller picture? Out of C [Long Beach] and D [Southern California], which one is smaller picture?
Male Student: C.
Female Student: C.
Teacher: It’s the smaller area. D is bigger picture. So even though, maybe 75% of the Iowans moved to Long Beach, there’s still other Iowans that live like in Newport Beach, or Anaheim, or other parts of the Los Angeles Area or even San Diego. So Southern California is more of a what answer?
Female Student: General.
Teacher: Yeah, and that’s what you want to defer to when you are kind of caught in between two, because, you guys, they’re making these tests for the masses. They are not making these tests for you. It’s for several thousand of kids so it’s going to be more general in nature. And remember, we’re not going to necessarily walk away with a lot of content from doing all of this. We’re going to be walking away with concepts. So when you get down to two choices and you have a big picture answer vs. a small picture answer, chances are it’s which one?
Female Student: The bigger picture.

And as the study wound down, it occurred to me that we were forging through the test-taking landscape, we were carving out concepts, things that would remain on the tests despite the ever-changing nature of the content. Such things we had discovered included understanding
the key words in the test question, reading up and down to understand the context of the
passage, scanning to get a quick recap, narrowing down the choices to two before
reexamining the question again and so forth. It was not rocket science but it had to be done,
and I was so sure from my uniform pedagogy that such skills would undoubtedly transfer.

But as the journals revealed, the students didn’t say much about being “enlightened”
by the intervention, and yet by the end, very few students said anything alarming either. By
the May journals, only 8 students of the 37 commented on their feelings about testing. The
others focused on the testing itself and did not offer their feelings about it. Advanced students
tended to speak in third person. Two of the three who commented said remarks like,
“Someone could have missed this question because they didn’t want to take anymore tests.
The person might try to put their testing fatigue aside and try harder.” There responses were
generally polite. Basic students spoke exclusively in the first person and said things like, “I
wasn’t feeling like testing and reading all of those paragraphs,” or “I was lazy that day,” or “I
missed the question because I was really hungry.” The test didn’t count in any way against
the students’ grade and they knew that. Finally, a couple of below basic students remarked, “I
guessed because I was lazy” or “I rushed threw the test.” Students would have to be
incredibly intrinsically motivated in order to keep up their high levels of effort. The tests not
“counting” could have been a factor that deterred them from trying their best. Yet for the
most part, by the end of the intervention, only about a quarter of the students attributed
laziness or lack of effort to undermining their performance. Most did not say either way.

In conclusion, although all groups revealed some degree of laziness or tiredness
affecting their performance, this was not for the majority of students; most did not make
comments about their feelings about the tests or the overall intervention. During the journal
entries, most students focused on the discussions and what they do if they could revisit certain test questions again. However, this may have been because the journal prompts did not directly ask them to state how they felt about testing. Rather, if a student felt that his or her testing performance was affected by laziness, tiredness, or some other factor affecting his/her motivation and the student mentioned it, it was considered in the efficacy section. Through the journals, only three students suggested that the intervention was especially helpful to them getting the “right” answer, two basic students and one below basic student.

**FOCUS STUDENTS’ PERCEPTIONS OF MOST AND LEAST EFFECTIVE ASPECTS OF THE Study**

As for the focus students, three of the six felt the actual think alouds were most useful, followed by two who felt that the classroom discussions were most effective, followed by one who thought taking the actual practice tests was most valuable. Three of the six focus students, one advanced, one basic, and one below basic, felt that the actual process of recording their thinking, of thinking aloud, was the most beneficial aspect of the intervention. As Rashawndra claimed:

> It helped me to clarify my thoughts more because I was relaying them to something . . . it helped me to break down my thoughts . . . I thought I was like maybe checking myself more and catching myself on things that I wouldn’t have ever noticed.

Reynaldo said, “because you think—you understand what you’re thinking instead of just in your head. You’re saying it so you understand more of what you’re thinking . . . . The think aloud is something good.” Finally, Jenaya stated, “Talking into the microphone [was the most helpful] because it helped me by like thinking about which questions that I chose and which ones that I didn’t chose.”
Two of the six, one advanced and one below basic, stated that the classroom discussions were most beneficial. Nathaniel said, “It [the discussion] helped me to see what other kind of thinking there is and how you could change yours if you ever need to” while Alejandro said, “The best part was listening or participating in the class discussions because I had more help, because people would answer the questions and they would say their thoughts so I would take those down too.” Finally, Bill, a proficient student thought that taking the actual practice tests helped him the most. He said, “I would say the practice test was the most helpful because I thought of a lot of different strategies to help me during that . . . .” Bill felt that the practice tests helped to reinforce the strategies that he was refining.

Concerning what was least effective, four of the six focus students felt that writing in the journals was least effective followed by two who thought that taking the practice tests was the least helpful aspect of the intervention. All of the advanced students and one basic student felt that writing in the journals was the least effective aspect of the intervention. For example, Nathaniel stated, “The least helpful to me was the journal writing. I didn’t get much improvement from it . . . . It had a purpose but I didn’t get to the next level because of it.” Bill thought that writing in the journals was least helpful because “it was a bit harder than doing anything else in the testing process.” Rashawndra shared this same sentiment regarding the difficulty of journal writing. She said:

It [the journal writing] was really hard for me, because a lot of the times, I can’t find the right words to put into my thoughts and I was trying like, to incorporate what we just learned into the discussions but sometimes I didn’t really know how to.
Reynaldo said, "[The journal writing] was boring... because I don’t care to write stuff down." These students seemed to struggle with the journal intervention because it was a challenge for them to capture their ideas in writing.

The less advanced test-takers, Alejandro and Jenaya, perceived taking the practice tests as the least important part of the intervention. For example, Alejandro said, "I didn’t really like taking the practice tests." Jenaya said, "I got better by taking the tests but the worst part was reading the articles." The less advanced students seemed to struggle with the practice tests because of the vast amounts of reading but they had trouble articulating why.

**STUDENT SUMMARY RELATIVE TO EFFICACY**

Through the intervention, students seemed to able to grasp a lot of complex constructs, but the degree to which they attained mastery appears inconclusive and within the course of this study, it seems that the intervention had little impact on their immediate test scores as outlined in the quantitative section. Many students, about one-third, stated that laziness or lack of interest was a factor in their test taking process at some point in their journals. A few students proclaimed that the intervention was really helping them to become more successful in getting answers correct.

Overall, five of the six focus students stated that either going through the process of thinking aloud or listening to and/or participating in the classroom discussions was the most effective part of the intervention. Several students commented that saying their ideas out loud helped them to think more clearly. Some also stated that listening to others thought processes gave them more ideas on how to approach specific test items. Only one student felt that taking the practice tests was the most valuable aspect.
Conversely, only the advanced and basic students claimed that the journal writing was least effective. Many expressed the difficulty of writing as the primary reason. Several didn’t find it enjoyable. Finally, both of the below basic students did not favor taking the practice tests mainly because of the difficulty in getting through the readings. Overall, it appears that the students felt that the least effective part of the intervention was the area that they had the most trouble with.
CHAPTER 5

INTRODUCTION TO THE CONCLUSION

This chapter reintroduces the initial problem, the purpose of the study, and the rationale behind it. It then investigates the research questions and explores the key findings followed by a discussion and limitations. Finally, it concludes with recommendations for instructional practice, future studies, and final remarks.

THE PROBLEM

With the advent of the federal law, *No Child Left Behind Act* (2001), standardized assessments are mandated in every public school across America. Since the federal government is now monitoring the success of schools all over the country, the way schools operate is radically changing yet there are still stark achievement gaps among those socioeconomically disadvantaged. Overall, African American and Hispanic children continue to lag behind (Haycock, 2006). In traditional high schools, test preparation is still seen as something that teachers have to do in addition to their curriculum, and it tends to occur some few weeks before a high-stakes test (Crocker, 2005; Durbin, 2002), far too late for many students who need to pass minimum competency tests like the CAHSEE to graduate from high school.

Historically, test preparation interventions give students insight about what a test will look like and what material will be covered on it, but such test preparation usually does not include in-depth analysis of specific test questions (Becker, 1990; Kulik et al., 1984; Samson,
1985), especially student-generated analysis. Clearly, there is a need for in-depth analysis of test items to aid students in becoming more reflective. And, as noted by Zimmerman (1995, 2000), self-evaluation permits students to become aware of their deficiencies so that they are able to engage in strategy refinement. For greatest effectiveness, development of good test-taking skills should systematic and thoughtful (Chapman, 2002). Such instruction should be a year-long activity incorporated into classroom instruction not a frenzied activity incorporated just a few weeks before a major assessment (Crocker, 2005; Jones et al., 2003).

**PURPOSE OF THE STUDY**

The purpose of this study was to see if a specific intervention could aid students in preparing for a high-stakes standardized test like the CAHSEE. The proposed intervention involved students taking several miniature Princeton Review practice CAHSEE tests. It then required that students explain their metacognitive approach (metacomprehension, problem-solving) to specific test questions during in-class discussions. Finally, it required students to write about (critical thinking) the impact that in-class discussions may have on their test-taking ability, particularly how their approach to specific questions might change in the future. In addition, six focus students participated in think aloud during the test-taking process and participated in interviews at the end of the study in which they explained which aspects of the intervention were least and most effective and why.

The intervention also explored if specific groups (advanced vs. basic, basic vs. below basic, advanced vs. below basic) improved at different rates over a 3-month period within a semester. The purpose of the focus students was to examine six students' test-taking processes closely. Half of these focus students were historically strong test-takers and half
weak. The final phase of the study included an interview with each of the focus students to investigate their perceptions about what aspects of the intervention were the most and least helpful.

**RATIONALE FOR METHOD**

This study was a mixed-methods action research study. While a quantitative approach captures scientific data of participants and may reduce the risk of researcher bias, it lacks the collaborative dialogue and participatory decision-making that a qualitative approach can bring (Denzin & Lincoln, 2005). Therefore, a study that includes both quantitative and qualitative data allows for maximum possibilities to not only report participant performance, but to investigate how to improve it, a fundamental aim of education.

**RESEARCH QUESTIONS**

The research questions driving this study were as follows:

1a. How, or in what ways, does a series of practice tests influence students’ readiness for the actual CAHSEE exam?

1b. How do student groups perform on average relative to each other (advanced vs. basic, basic vs. below basic, advanced vs. below basic)?

1c. How does practice test achievement between more advanced and less advanced focus students compare?

2a. How, or in what ways, do a series of classroom discussions influence students’ readiness for the actual CAHSEE?

2b. Specifically, what verbal articulation and efficacy patterns can be observed in the focus students?

3a. How, or in what ways, do a series of journal reflections influence students’ readiness for the actual CAHSEE?
3b. Specifically, what written articulation and efficacy patterns can be observed in the student groups (advanced, basic, below basic)?

4a. How, or in what ways, do a series of think alouds influence students’ readiness for the actual CAHSEE?

4b. Which aspects of the intervention do the focus students perceive to be most and least effective and why?

**KEY FINDINGS**

A question that emerges through the interventions offered in this study is, “Can an intervention specifically designed to reveal students’ thinking processes prepare them for tests like The California High School Exit Exam?” If one looks at the results of this study literally with a quantitative eye, the answer may be “no” for the basic students who did not make any gains, or “marginally” for the advanced students who gained approximately .75 of a point by the end of the intervention and “possibly” for the below basic students who gained 1.11 points between April and May.

On average, the students in this study did not improve much quantitatively and yet, as Becker (1990) found in her meta-analysis of 23 SAT studies, for every 10 hours of coaching, students only improved between 1 and 6 points. It takes many hours of coaching students in test preparation to show gains and the gains, on average, are minimal. If one looks only at the numbers in this study, the story stops there. However, if one looks more deeply into the thinking behind the items, it’s easy to see that students made gains in their thinking and that the work of future studies could lead to more gains.

Three main qualitative findings emerged from this study that reach well beyond the students studied into the instructional world of educators. The first is that students of all levels (advanced, basic, below basic) had many challenges when approaching items, and by
the end, many still struggled, particularly with revisiting the text. The second finding that emerged through the focus student data is that the behaviors of high and low test-takers appear to be genuinely different; high test-takers tend to approach standardized test questions in a cyclical fashion applying various strategies in different orders, sometimes multiple times. Low test-takers, on the other hand, tend to approach such questions in a lock step linear pattern without altering their approach even if certain items types warrant such change. Finally, the third finding is that students' feelings about testing may play a role in their performance and that these feelings may be beyond the educator's control.

**THE IMPORTANCE OF EXAMINING THE WAY STUDENTS THINK AND PROCESS**

Although an enormous body of literature has focused on standardized test scores in the past few years (Allensworth, 2005; Borg et al., 2007; Causey-Bush, 2005; Cizek, 2002; Craig, 2004; Darling-Hammond, 2006; Garcia, 2003; Haney, 2000; Haycock, 2006; Hong et al., 2006; Jones et al., 2003; Kohn, 2000; Perkins-Gough, 2005; Plake, 2005; Ruth, 2001; Sacks, 1999; Watanabe, 2007), little, if any, research has investigated the way K-12 students think while they test. This is a crucial area that has the capacity to give educators tremendous insight on how to develop specific strategies to assist students. It should not be underestimated.

Standardized test scores, by themselves, do not give teachers enough information. While English language arts tests may indicate areas of strength or weaknesses related to strands such as vocabulary knowledge, reading comprehension or writing strategies, they are general in nature and do not reveal the specific, intricate processing that students experience as they select answers. In addition, although meta-analysis studies have been conducted
examining student test scores (Becker, 1990; Kulik et al., 1984; Samson, 1985), none of these analyses have documented student processes while they test or even, while they review items. It is an area that is not well researched, and, this information, this researcher argues, is a precursor in order to help students to improve their test scores. The skills that students experience weaknesses in must be addressed instructionally and such instruction should not be confused with “teaching to the test.”

Evidence from this study indicates that the students in this study had many difficulties processing test items. These difficulties took multiple forms including how to interpret a test question, how to recognize key words, how to select the most correct answer, and how to revisit the text. Although several researchers suggest strategies to use when assisting students with standardized tests (Becker, 1990; Crocker, 2005; Kitsantas; 2000; McClain, 1983; Pressley et al., 1997), many of the strategies are really quite general. For example, one of the strategies Crocker (2005) suggests is helping students to learn how to determine what a question is really asking.

**Interpreting the Test Question**

As the teacher researcher in this study learned, “how to interpret a test question” seems very straightforward until students are faced with the challenge of putting a question into their own words. Early in the study, when I asked advanced students to put the test question, “Which of the following sources would give the best information on northern lights mythology?” into their own words, I got responses like, “Use the question to base it on yourself,” “Read the whole question and make sure you know—like have some past knowledge about the question,” or “Use the vocabulary in both the question and the answers
to figure out which one works best.” The students did not seem to understand how to put the question on northern lights mythology into their own words. They didn’t recognize that this was a process in which they had to put the question into their own language, a language that was comfortable and made sense to them. It was as though they had never been given permission to do this before. Many seemed to think it was more complex activity than it actually was. But until I knew how they understood the strategy, I could not help them.

**Recognizing Key Words and “Correct” Answers**

Recognizing key words was another area in which the students revealed that they needed practice. By the end of the study, many students learned to pay attention to contrasting terms like most and least. For example, one focus student stated during the May think aloud, “I kinda think that all of these [answers] could be answered by further research. But the key word was best because it’s bold and capitalized.” In addition to recognizing key words in the stem of the test question, as students eliminated answers, they were also introduced to the challenges of differentiating between a correct answer and the most correct answer. The awareness that there was more than one correct answer was new concept to some of them, even the advanced students. For example, regarding a question about what she learned, an advanced focus student stated in an interview, “... the method where you cross out wrong answers and what you [the teacher researcher] said about how it could be two answers, but which one best suits, like I never really thought about that.” Like a below basic student who seemed unaware of how to eliminate all answers until it was pointed out, this advanced student seemed unaware that there could be more than one correct answer.
Revisiting the Text

Finally, revisiting the text was an area that all groups seemed to need the most practice on. The decision of how to revisit the text was one that students of all levels struggled with up until the end of the intervention. They often did not know how to approach the text. For example, many wrote in their journals that revisiting the text meant to skim and scan. But they did not take specific test questions into consideration. For example, if the test question asked for the meaning of a sentence, the idea of scanning wasn’t going to help them zero in. They needed to read up and down, above and below, the sentence in question to understand the context. On other items, asking for the author’s tone or the author’s perspective, they had to learn to limit their own personal attitudes and opinions in order to not be blinded by their own perspectives.

THE DIFFERENCES BETWEEN HIGH AND LOW TEST-TAKERS: CYCLICAL VS. LINEAR

The second finding that emerged from this study reveals that effective standardized test-takers tend to utilize strategies in a cyclical fashion; they don’t follow a lock step, linear approach. Rather, they revisit the question, text, and answer choices as many times as they need to and in varying orders.

The most accurate test-taker, Nathaniel, demonstrated interesting patterns that revisited strategies continually. For example, he would often read the test question and then take time to interpret it. Then, he tended to extract key words in the test question that helped him to realize where he would go in the text. On a question asking, “Instead of following a strict vegetarian diet, the author of the second article encourages students to

_____________,” Nathaniel said, “To answer this, I will go back to the article into the
second argument.” He paid attention to the question and noticed it asked about the second article and not the first. Then he stated, “And um, I would look in the last paragraph in the conclusion where the author would sum up . . .” Because the question asked about what the author encouraged, Nathaniel thought that this information would be summed up in the last paragraph. He had an automatic sense of where to find either confirmation or non-confirmation of the choices.

However, when the last paragraph didn’t tell him everything he needed to know, Nathaniel refined his approach, continuing to actively engage while he searched. He added, “So if your position is to find out what the perfect diet is . . . there’s the fourth sentence in the second paragraph . . .” Finally, when he thought he recognized the answer through textual evidence, he would then go back and refute the other possible answers just to make absolute sure he was correct. And when he explained his thought processes, they matched with what he actually did, unlike the below basic test-takers Alejandro and Jenaya, who claimed to do things like revisit the text when their think aloud transcripts showed no evidence that they had done so.

Students like Nathaniel could be used to illustrate the complex thinking and versatile navigation skills that high level test-takers experience. At least this way, students would have a model in which they could compare their own processes to. However, this research does not attempt to formulate an ideal way to build an effective test taker but rather it reveals the complex nature of all students educators serve and shows that we should provide inquiry-based instruction in which students can sample a plethora of effective ways to approach test items. Students should feel empowered to experiment with their own processes in order to feel ownership of them.
THE FEELINGS THAT STUDENTS HAVE REGARDING TEST-TAKING

The last finding involves the importance of recognizing the feelings that students have regarding testing. Journal evidence from this study suggests that approximately one-third of each of the advanced, basic, and below basic groups attributed laziness or tiredness as factors affecting their performance. Although much literature suggests that low achieving students are less motivated and less self-reflective than their higher achieving peers (Hong et al., 2006; Kim & Goetz, 1993; Kitsantas, 2002; McClain, 1983; Zimmerman, 2000), evidence from this study suggests that individuals in all groups experience a lack of motivation concerning standardized testing. Yet this brings about a new question: Why did the lowest performing group improve the most? Did this improvement have anything to do with a change in their feelings?

Putting It All Together: What Made the Lowest Performing Group Improve the Most?

Evidence from this study suggests that the lowest performing students, the below basic test-takers, improved the most. This is contrary to earlier research, which suggests, that typically when coached, advanced students will improve more than less advanced students (Kulik et al., 1990; Vernon, 1954). However, the answer to the below basic test-takers’ improvement in this study may be from a few factors. The first may be that because the below basic test-takers were more homogenous as a group than their basic counterparts, they had more opportunity to improve. They were more likely to be in their Zone of Proximal Development (Vygotsky, 1978) because the range of their scores was so much closer than the basic group. The other reason may be that they benefited more from being reassured than
their higher-achieving peers. As observed early in the study, the below basic test-takers seemed to have a lot of bad feelings around test-taking. They made statements like, “Why do we have to take a test?” “I don’t know how to read,” and “It’s the fact that it’s a test.”

Early on, there was a stark resistance in the below basic group that I did not see in the other groups, and I found myself continually saying things to reassure them like “It’s so important to take a few minutes to think about how we get caught up in the wrong things when we’re looking at these tests, when really, they are manageable if we just break it down.” The low achieving students seemed to enter the study with the lowest skills and the worst feelings about tests so it is not surprising that they would improve the most. It is important, when considering low test-takers, not only to teach them test-wiseness skills, but to reinforce their ability to succeed. Many are fragile and have bad feelings around testing.

**The Focus Students’ Feelings and Preferences**

Regardless of the overall class responses, some focus students appeared to gain confidence through the intervention. By the end of the think alouds, Rashawendra pressed through her fatigue and Bill made his decisions with more authority, citing specific information from the text. However, Alejandro and Jenaya seemed to think that they had improved throughout the intervention when their test scores did not reflect this. Alejandro said that the reading the selections “wasn’t really hard” and that in order to figure out the answers, he “went back to the text” and “went through the answers and crossed out the ones that didn’t make sense.” He claimed “to get the answers down to two” before he eliminated; however, a number of his think aloud transcripts do not show this. Jenaya said, “I think I got better by taking the test,” “I do like better in testing now, I think so,” yet her scores went
down to only 8 out of 25 correct by the end of the intervention. She had the lowest test scores in the study. Jenaya, like Alejandro, looked at the questions first, and then read the excerpt looking for the answers to that information. She recommended other test-takers to “read the questions first, read the article and eliminate answers.” This approach of reading the question first may not be best for low test-takers who may search for the answer too quickly, without first understanding the overall concept of the reading.

This idea of low achievers being inaccurate about their processes is consistent with the research of Hart (1965, as cited in NRC, 2001), McClain (1983), Kim and Goetz (1993), and Hong et al. (2006) regarding how higher achievers have stronger, more realistic feelings of knowing than that of their less able peers. However, the below basic test-takers seemed to feel more hope about being successful, even if their perceptions were off. As Alejandro noted, he felt more hopeful about being successful and he considered the test-taking intervention as “a great process to do.” Jenaya seemed to feel more successful as she stated, “I seen a pattern, like one month I did bad, the other month I kind of did good, and the other month I did terrific.”

In looking at the focus students’ preferences, three valued the think aloud process the most because the nature of recording their thoughts caused them to think more carefully. As one high achieving focus student put it, “I was more careful on the think alouds because I didn’t want to mess up. I don’t like being wrong.” Another less motivated basic student stated, “I only answered the think alouds. On the other questions, I guessed on them . . . . On the think alouds, I just tried my best to answer them—to get the answers right . . . they were being recorded.” Interestingly, the act of recording one’s voice seemed to motivate students into being more careful and precise about what they said. Two others valued the nature of
discussions the most because they could listen to how each other thought and processed. Two found it most interesting to see how their peers thought. One, Alejandro, a below basic student, stated that, “The best part was listening or participating in the class discussions because I had more help, because people would answer the questions and they would say their thoughts so I would take those down too.”

Although the focus students’ preferences for the most effective aspect of the intervention were mixed, their feelings about the least effective aspect of the intervention were more distinguished. For example, the basic, proficient, and advanced focus students felt that the journal responses were the least helpful while the below basic and far below basic students felt that the practice tests themselves were least helpful. The basic and above students resisted writing about the intervention while the below basic test takers resisted reading the most. This suggests, along with the literature, that low achievers have the hardest time reading (American Institutes for Research, 2005).

**DISCUSSION**

The results of this study go counter to the literature when examining that all high achievers are motivated and all below basic students are not. In fact, about one-third of the students in each class struggled with motivation in all classes. Many had tired of testing and they did not want to continue. One of the proficient focus students indicated that he didn’t think it was fair that he had to take another test after taking the practice CAHSEE district test in March, and the CST in May. He felt over-tested. And yet ironically, he felt that the most useful aspect of the study was taking the practice tests because he could practice the strategies that he was learning.
Not All Low Achievers Lack Motivation

Contrary to the literature, evidence from this study also suggests that the two of the three low-achieving focus students were actually quite motivated to improve. Alejandro said, “I say anything I have to [while thinking aloud] because I’m trying to learn from what I have to do… I tried to use different strategies.” Although Jenaya didn’t say much about wanting to improve possibly because she struggled so much with the language, she believed her efforts paid off. During her end of study interview, she said, “I do better testing now, like the reading.” Clearly, these focus students wanted to get better at testing.

However, despite what they wanted to happen, they may not have improved much because of their lack of skills. As other literature attests (Kitsantas, 2002; Pintrich & De Groot, 1990; Schunk, 1989; Zimmerman, 1989, 1990; Zimmerman & Martinez-Pons, 1986), these low achieving students were not able to self-regulate as well as their higher achieving peers. They were not as aware of what they did while they tested and their descriptions of their behaviors did not match the evidence.

Lack of Motivation May Be Beyond the Teacher’s Control

Students, for the most part, didn’t outwardly say that they appreciated the intervention and felt like it made a positive difference. To the contrary, about one-third of the students commented on their feelings about testing and most of these comments suggested the test preparation was something that they didn’t want to give their best effort to. Some of these students were quite vocal like Reynaldo who felt that tests were “pointless.” Most others indicated laziness or a sense of tiredness. This information is important because if students are feeling resistant, then instructors, like myself, should be cognizant of discussing why they
feel this way. In addition, instructors should not be blamed exclusively if students’ standardized test scores, on average, do not rise. Other factors may be at work, factors beyond the educator’s control.

Teaching Test Wiseness Is Not Teaching to the Test

In her analysis of the Texas state standardized testing system, Linda McNeil (2000) discovered that teaching to test narrow the possibilities of the curriculum greatly and puts extraneous limits on what teachers and students can achieve. However, it is important to clarify that test-wiseness, as defined by Milman et al. (1965), is largely interdependent of the examinee’s knowledge of the subject matter for which the item supposedly measures. The purpose of preparing students for tests, particularly English language arts tests that have a preponderance of new reading material, should be to develop their test-wiseness skills as they approach such reading material (Sarnacki, 1979; Smith, 1991). This is not teaching to the test but rather teaching test-takers to be wise as they approach items.

The Way Test Scores Are Reported May Make a Difference

Evidence from this study indicates that many of the focus students were, in fact, very capable test takers who happened to be African American and Hispanic. The California Department of Education CAHSEE data portrays students by subgroup with minority children of African American and Latino backgrounds at or near the bottom compared to their white and Asian peers. Yet if we, as a profession, are going to improve test scores for these students, then perhaps we should disperse the data differently, showing who is performing well within the African American and Hispanic populations. Evidence from this
study demonstrates that these students are out there. They do exist. In addition, evidence from this study also suggests that the students who performed below basic and far below basic were still motivated to improve and they were appreciative of the intervention. They wanted to become better test-takers. They simply didn’t have the skill sets.

**LIMITATIONS**

Several limitations are apparent in this study. However, the sample size, the reconfigured tests, the time of year during which the study took place, and the length of the study appear to be the most glaring.

The sample size investigated, 37 total (classes of 11, 17, and 9), was small and made it increasingly difficult to produce accurate, quantitative data. As one focus student indicated, he only tried on the focus questions because they were being recorded. On the others, he guessed. He was in a class of nine so his style of guessing could have impacted the others’ results in the class, those who were really trying. In addition, the number of focus students, only six, can’t be a representative sample of all students. However, it was amazing that six students could show such a vast range of skills across the spectrum and it is a stark reminder that students’ processes are unique to each individual.

In addition to the size of the study, there were problems with the reconfigured Princeton Review tests. Even though the tests appeared equal because they were broken down systematically through similar reading categories, as seen through the evidence, all groups performed the worst on Test B, regardless of when they took it. In addition, all performed the best on Test A. Regardless of the effort to create equal tests, they did not appear to be.
In addition to the sample size and the reconfigured Princeton Review tests, the time during which the study was conducted may have influenced the results. The study was during a 3-month window toward the end of the year, during the months of March, April, and May. Not only was the school year starting to wind down, students were faced with other tests like the district mandated practice CAHSEE in March and the CST in May. Students completed a lot of tests in a short time and some students felt that the intervention I was doing was less crucial than the district and state tests. One stated, "I really didn’t feel like doing the test. I had other stuff on my mind like other schoolwork and I kind of got lazy . . . I really had my mind set on the STAR testing." He didn’t appear to understand that the intervention was designed to improve his test-taking abilities on other tests like the STAR.

In addition to the hectic time of year in which the study took place, the length of the study may have been too short. In her study of 23 SAT studies, Becker (1990) found that for every 10 hours coached, students improved on average 1 to 6 points on the SAT. Kulik et al. (1984) found that students needed to be coached 9 hours or more to show the greatest effect size while Messick and Jungeblut (1981) found that students needed at least 3 hours of coaching to show gains. In this particular study, students were coached for about 3 hours (six 30-minute discussion sessions).

**Recommendations for Instructional Practice**

One way to gauge students’ perception of their test-taking skills and feelings is through the use of a rubric similar to the spectrum of behaviors developed in this study (see Table 14). Such a tool could be used to have students self-evaluate their test-taking skills so they can actually measure and mark their progress over time. Such methods like rubrics that
invite self-evaluation and self-reflection have been known to yield student achievement
(Marzano, Pickering, & Pollock, 2001; Schunk, 1989; Strong, Silver, & Perii, 2001). This
approach might make students feel less put off by test-taking interventions and more aware of
how they can change. Such an approach could help make teachers aware of how students are
feeling with regard to standardized tests.

In addition to having students analyze their testing behavior relative to a rubric, it
might be beneficial for teachers to teach specific test-taking skills within the curriculum. For
example, teachers could assist students more in the process of interpreting words and phrases.
In the beginning of this study, I assumed that students understood what I meant when I asked
them to interpret the test question into their own words. Conversely, they touted back phrases
that they seemed to think I wanted to hear. They didn’t seem to understand how interpreting
is a personal process in which words and phrases must be transformed into their own
language in order to have real meaning. This process of interpreting is powerful and students
need ample practice putting academic language into their own words.

In addition to practicing interpreting, teachers need to give students opportunities to
eliminate incorrect choices in order to improve their sense of logic. For example, in the
general curriculum, while teaching, teachers could give students prompts such as “Is the Matsuri
Festival year-round or annual?” Such questions could help the students to
make a decision and back up their choice with evidence from whatever they are reading. The
students in this study struggled to tie their critical thinking directly to evidence from the text.
They need a lot of practice in order to do so. In some instances of instruction, students may
need prompts presented by the teacher in which they choose either “this” or “that” and
provide appropriate rationale as to why.
Finally, students need a lot of practice applying reading strategies appropriately. As seen in this study, many students applied reading strategies ineffectively. For example, when a test item asked students to interpret a sentence from the text, many students felt that an appropriate solution would be to skim and scan. This more summative, general strategy did not effectively fit the search for specific knowledge at the sentence level. Students appear to need a lot of practice understanding the scope of the test items at the word, sentence, paragraph, and excerpt levels. They need to learn to differentiate certain strategies at varying levels. As this study revealed, students needed the most practice in this area, possibly because it is the most complex with the least set of consistent rules; students often need to make decisions intuitively. They need opportunities to develop such skills while they read and to have a knowledgeable adult with which to share their confusion when adopting specific strategies. They need to understand that determining appropriate reading strategies can be a "messy" process.

In addition to having students practice various reading and writing strategies, teachers should encourage students to think aloud during activities in order to enhance their speaking and listening abilities. For example, in a paired reading activity, one partner could read a paragraph of text to another partner and the receiving partner could explain what he or she thinks the paragraph means. This idea of sharing information with each other verbally has the power to enhance students' metacognition and may help students improve their speaking and listening skills.

In addition to what teachers can do to help students, much of the information gleaned from this study can assist teachers in staff development environments. For example, prototypes of test-takers all across the board could be developed based on think aloud data.
Such prototypes could be used during staff development so teachers could discuss which strategies they could use with which prototypes. For example, Rashawndra might be ready to distinguish between the most correct answer while Alejandro may need to first master how to interpret the test question into his own words. This could also be a way to enrich teachers’ knowledge about reading strategies since many non-English language arts teachers may simply not know what to do, specifically, in order to assist students. Looking at prototypes of students and asking teachers what strategic interventions they would use might make the act of reflecting on student standardized test performances more enjoyable for teachers. It could also expand many of their understandings well beyond analyzing test scores.

**RECOMMENDATIONS FOR FUTURE RESEARCH**

As the evidence from the classroom discussions showed, all groups struggled with knowing how to match a particular reading strategy with a certain kind of item. For example, students found it hard to recognize that a question asking students to interpret at the sentence level might require that they revisit the text and read above and below the sentence in question. Future studies that pre-aligned many similar item types with text would give the chance for students to master complex reading skills because they would be exposed to the same strategy multiple times. Because the nature of the Princeton Review Practice CAHSEE tests demanded that the students continually switch their reading approaches depending on the test items, it made it difficult for many of them to master the reading skills needed. A study that took this into account could possibly yield more promising results on particular reading items. Becker (1990) refers to such items as “novel item types.”
As the evidence from the journal and focus student data showed, many of the students in this study, about one-third of all of the groups, were affected by feelings of laziness or tiredness. This is a crucial finding because such a lack of motivation could skew data sets into a downward direction despite what the teacher attempts to do. If I had been a typical teacher in my district with my name attached to students’ progress on tests, I could easily be accused of not teaching my students anything. In fact, if I worked in some districts, I might even be denied merit pay because my students’ scores didn’t rise quickly enough. It is important to investigate why certain students experience these feelings and not jump to conclusions that it is automatically due to a lack of instruction.

Finally, it might be interesting to conduct future think aloud studies to see if the act of thinking aloud improves students’ outlooks about testing. Many focus students thought it was the most useful part of the intervention. Many seemed to share an allegiance to the audio-recorder. As Reynaldo attested, he tried on the think aloud items but guessed on everything else. It would be important to see if this act of being recorded made other students more motivated to perform.

**Final Remarks**

Although there has been a great focus on standardized testing currently, too much value has been placed on test scores when there is actually so much teaching and learning opportunity that resides in examining the ways that students come up with “the answer.” To weigh so much on test scores alone and not to put any weight into how students are processing seems like a missed opportunity.
Alphie Kohn (2000) contends that standardized testing actually creates a *double whammy* for low socioeconomic students since such test preparation not only robs students of precious instructional time, but sets up a situation in which teachers use a drill-and-kill approach. Yet such an approach is the result of irresponsible test preparation, instruction that is haphazard and not well thought out. If, on the contrary, instruction is based on student weaknesses and if it addresses those weaknesses in specific ways that make students more reflective thinkers, then it seems entirely justified. Such learning goes beyond the test.

Through the intervention, the students in the discussions revealed a lack of skill in many ways, and even after instruction, these ways were not magically fixed. It takes a lot of instruction and practice to change their skills and such instruction should be imbedded in the regular curriculum (Crocker, 2005) or at least introduced gradually throughout the calendar year so test preparation does not dominate the instructional landscape. However, one of the interesting findings in this study is that instruction can actually be created in reaction to how students process on standardized tests. In other words, the instruction is a direct response to student processes. As educators, we have to be sure to differentiate between haphazard, shallow test preparation instruction, and in-depth test preparation instruction that is in direct response to student skill levels. While the goal of standardized testing is to often measure critical thinking skills, ironically, such a measure is hard to improve if one only examines test scores and does not investigate the thinking processes that students experience while they test.

As this study shows, whether examining the focus student data, the classroom discussions, journal responses, or focus student interviews, how students were thinking before they made their decisions gave this teacher researcher critical insight. Without that
information, it would have been very difficult to assist students with specific strategies. This is an area that served as a necessary precursor that allowed this teacher researcher to refine her strategic approach. And even so, much work apparently still needed to be done, particularly with revisiting the text. Such test-taking strategies are highly thought provoking and should not be confused with rudimentary approaches of teaching to the test.

In addition to how students are thinking, it is also important to understand how students are feeling. Is the intervention targeted to meet their needs? Do they feel competent? Do they feel motivated? If students aren’t buying into the mandates and feeling the need to intrinsically try, it is impossible to force them to do their best. However, it is important to stop and discuss the feelings that this standardized testing culture creates. It is important for the students to see the teacher as an ally who is on their side trying to help them to improve. It is important for students to see themselves as a necessary part of that process of improvement. Their insight is critical.

Finally, it is imperative to capitalize on what strong test-takers, like Nathaniel, in this study, do when they test. Such test-takers could serve as models for other students. However, it is important to note that each learner will take a slightly different path. In general, the more successful test-takers in this study demonstrated a cyclical approach to test questions, in which they addressed certain themes multiple times and in varying orders. They did not stay in a lock step design that was linear in nature. Once given appropriate strategies, students need to be reminded that while testing, they have the freedom to choose their own personal path; they have the power to make it a creative process. In other words, the tests they may be taking are standardized, but they are not!
REFERENCES


APPENDIX A

PROTECTION/ANONYMITY OF SUBJECTS:

PARENT PERMISSIONS LETTER
San Diego State University

Parent Permission Form for Child to Participate in Research

Can A Constructivist Approach Be Useful To Prepare Students for Standardized Tests?
How Ninth Grade English Language Arts Students Prepared For
The California High School Exit Exam (CAHSEE).

Your child is being asked to participate in a research study. Before you give your consent for your child to volunteer, it is important that you read the following information and ask as many questions as necessary to be sure you understand what your child will be asked to do.

Dear Parent(s):

As a doctoral student in the San Diego State University and The University of San Diego Joint-Doctoral Program, I am interested in conducting a study about how ninth graders think when faced with English Language Arts standardized test questions in preparation for the California High School Exit Exam (CAHSEE). Dr. Barbara Moss, SDSU Professor of Education (PhD), Dr. Marcie Bober-Michel, SDSU Professor of Educational Technology (PhD), and Dr. Heather Lattimer, USD Professor of Education (EdD) will be supervising this study.

Purpose of the Study: The study is designed to determine if a new approach to test preparation will help ninth graders effectively prepare for the English Language Arts portion of the California High School Exit Exam (CAHSEE). All of my freshmen students will be invited to participate in this research study.

Description of the Study: I am asking that you give permission for your child to take part in this research study. Although all students will take practice tests, listen to other students talk about certain test questions, and write journal reflections as regular class assignments, those who participate in the research study may be asked to talk about how they tried to solve certain test questions during discussions and they will be asked to allow me to analyze their test scores, their comments during discussions, and their journal reflections.

During class, your child will be asked to:

- Take three, 25 question Princeton Review Tests in preparation for the English Language Arts CAHSEE in March, April and May. Each test takes about forty-five minutes to an hour to complete. **All students will do this as a required classroom assignment.**
- Review score distributions showing how well participants in the class did on test items (individual student names will not be included). This will take only a few minutes. **Only students who decide to participate will have their test data included.**
- Discuss (and/or listen to) student approaches for solving specific tests questions such as eliminating certain answers, analyzing key words, and double-checking responses with a text. These specific test questions will be selected by participants for review and they will be audiotaped for teacher analysis. This part will take twenty to thirty minutes and will
occur on a different day than the day of the test. All students will listen to discussions but only those who are participating will comment and be audiotaped.

- Write in a journal about how the class discussion may have influenced his/her thinking on future test items. This part will take ten to fifteen minutes and it will occur after the whole class discussion on certain test questions. All students will do this but only those who participate will have their journal responses analyzed by me.

- Some students may also be asked to volunteer to be focus students. This means that these students will also participate in think alouds in which they test in a private setting (the library) and explain what they are thinking about certain test questions into an audiotape recorder while they test. In between tests, all focus students will be shown transcripts (typed words of what students said during think alouds) to see if they can recognize the behaviors of strong test-takers. At the end of the study, focus students will be interviewed to see which strategies they think were the most helpful and least helpful for them to prepare for the CAHSEE. If you decide you want your child to participate as a focus student, please understand that your child may not be selected due to the small number of students needed for this part of the study.

If you agree to have your child participate in the study, please understand that his/her test scores will be used to see how groups of students performed over time on practice tests. However, results of your child’s individual scores will not be used. I am interested in looking at how groups do, not individual students.

If you agree to allow your child to participate in the study and your child speaks during class discussions, please understand that your child’s voice will be recorded for analysis. Such analysis may include looking at how your child tried to solve a test problem or things that he/she did incorrectly like choosing the wrong answer because he/she misread or did not understand the test question. This may be necessary to help students learn how to form correct responses on future test questions.

If you agree to allow your child to participate in the study, your child’s journal writing will be analyzed. Such analysis may include comparing your child’s response about a test question with other classmates’ responses to check for similarities and differences.

Risks: Risks associated with the study are that some students might feel stressed when taking practice tests. This risk is managed by giving students three opportunities to improve their scores. In between each test opportunity, students will either participate in or listen to, how students solved certain test questions in order to become better test-takers. Some students might feel embarrassment about their own test score results. This risk is managed by making sure that no one knows anyone else’s test results. However, instead of focusing on a student’s individual scores, we’ll focus on how the class did on test questions. Student names will not be shared, only the last four digits of a student’s student ID number will be used so a student can see how he/she did but no one else can tell how he/she did. In addition, all of the activities will take place during class and a student’s performance will not affect his/her grade in any way.

Benefits. A student may benefit from participating by learning what types of test questions will be on the English language arts section of the CAHSEE as well as learning strategies to solve
certain test questions such as analyzing the stem of test questions, recognizing key words, and eliminating wrong answers. However, I cannot guarantee that any child will receive any benefit from participating in this study. The benefits to scientific knowledge that may result are that educators may learn how students think when approaching certain test questions and that may impact how teachers teach test preparation in the future. Educators may also discover ways to help students throughout California prepare for and pass the English Language Arts section of the CAHSEE.

Confidentiality: In order to protect your child’s confidentiality, the data for this study that contains student names will be stored on my personal, password-protected computer at my home location. All signed consent forms and paper test data will be stored in my classroom in a locked file cabinet. Recordings of classroom discussions will also be kept in the same locked file cabinet. I, the teacher/researcher, will be the primary one who has access to this information although I will hire a professional service to transcribe the classroom discussion data. In addition, my university chair and committee members may also see the data during discussions of how to best analyze the material. Upon completion of the study, all test data will be kept in a locked file cabinet by me, the teacher/researcher, for three years. After that point, all files, both electronic and paper, will be destroyed.

Voluntary Nature of Participation: Permission for your child to participate in this study is voluntary. Your choice of whether or not to allow your child to participate will not influence your future relations with San Diego State University, The University of California, or The San Diego Unified School District. If you decide to allow your child to participate, you are free to withdraw your consent and to stop your child’s participation at any time without any penalty.

Questions about the Study: If you have any questions about the research now, please ask. If you have questions later about the research, you may contact myself, Mary O’Shea at 858-496-8370 X4513 or via email at moshea@sandi.net. If you have any questions about your rights as a participant in this study, you may contact the Institutional Review Board at San Diego State University (telephone: 619-594-6622; email: irb@mail.sdsu.edu).

Consent to Participate: The San Diego State University Institutional Review Board has approved this consent form, as signified by the Board’s stamp. The consent form must be reviewed annually and expires on the date indicated on the stamp. Your signature on the following page indicates that you have read the information in this document and have had a chance to ask any questions that you have about the study. You have been given a copy of this consent form. By signing this consent form, you are not giving up your legal rights.

If you choose to allow your child to participate in this study, you can change your mind and withdraw consent to allow your child to participate at any time. If you decide not to allow your child to participate now, he/she cannot be added as a participant once the study begins. No matter what choice you make, your child’s grade in the class will not be affected in any way.
Please check the box and/or boxes that you give permission for:

☐ I would allow my child to participate as a participant in the CAHSEE study. I give consent for Mrs. O’Shea to analyze my child’s test score results, his/her comments during class discussions, and his/her responses during journal reflections.

☐ I would not allow my child to participate as a participant in the CAHSEE study. Even though my child will take practice tests, listen to certain test questions during class, and write reflections in his/her test journal, I do not want my child’s test data used or his/her discussion comments and journal entries analyzed.

☐ I would allow my child to participate as a focus student in the CAHSEE study.

First Name of Child (please print) __________________________ Last Name of Child (please print) __________________________

Name of Parent/Guardian of Participant (please print) __________________________ Date __________________________

Signature of Parent/Guardian of Participant __________________________ Date __________________________

Signature of Investigator __________________________ Date __________________________
APPENDIX B

PROTECTION/ANONYMITY OF SUBJECTS:

STUDENT PERMISSIONS LETTER
San Diego State University

Assent to Participate in Research as a Participant

Can A Constructivist Approach Be Useful To Prepare Students for Standardized Tests?
How Ninth Grade English Language Arts Students Prepared For
The California High School Exit Exam (CAHSEE).

Dear Student:

As you know, my name is Mrs. O’Shea and I am your English/Public Speaking teacher. I am trying to learn more about how students think and respond when faced with standardized test questions. I am interested in studying you and your classmates to see if giving you test preparation a few times per month for many months and discussing and writing about certain test items will help you prepare for the California High School Exit Exam (CAHSEE).

I am asking you to take part in a research study. Although all students will take practice tests, listen to other students talk about certain test questions, and write journal reflections as part of regular class assignments, those who participate in the research study may be asked to talk about how they tried to solve certain test questions during discussions and they will be asked to allow me to analyze their test scores, their comments during discussions, and their journal reflections. I want to investigate what strategies will help you to perform more effectively on standardized tests. I am trying to learn more about what is helpful to prepare students for tests like the CAHSEE.

During this study, you will be asked to:

- Take three, 25 question Princeton Review Tests in preparation for the English Language Arts CAHSEE in March, April and May. Each test takes about forty-five minutes to an hour to complete. All students will do this as a required class assignment.
- Review score distributions showing how well participants in the class did on test items (individual student names will not be included). This will take only a few minutes. Only students who participate will have their data included.
- Discuss (and/or listen to) student approaches for solving specific tests questions such as eliminating certain answers, analyzing key words, and double-checking responses with a text. These specific test questions will be selected by participants for review and they will be audiotaped for teacher analysis. This part will take twenty to thirty minutes and will occur on a different day than the day of the test. All students will listen to discussions but only those who decide to participate will comment and be audiotaped.
- Write in a journal about how the class discussion may have influenced your thinking on future test items. This part will take ten to fifteen minutes and it will occur after the whole class discussion on certain test questions. All students will do this but only those who participate will have their journal responses analyzed by me.

If you agree to participate in the study, your test scores will be used to see how you and your classmates performed over time on practice tests. Results of your individual scores will not be used. I am interested in looking at how groups do, not individual students.
If you agree to participate in the study and you speak aloud to the class during discussions, your voice will be recorded for analysis. Such analysis may include looking at how you tried to solve a test problem or things that you did incorrectly like choosing the wrong answer because you misread or did not understand the test question. This may be necessary to help students learn how to form correct responses on future test questions.

If you agree to participate in the study, your journal writing will be analyzed. Such analysis may include comparing your response about a test question with other classmates’ responses to check for similarities and differences.

Voluntary Participation: You are the one to decide to take part as a participant in this study. All of the activities (the practice CAHSEE tests, the test question discussions, the journal reflections) will take place during class and your performance will not affect your grade in any way. I will not be upset with you if you decide not to share your test scores or your thoughts during discussions or your journal writing for analysis. If you do decide to take part in the study, you may change your mind and stop participating any time you want.

Risks: Risks associated with the study are that some students might feel stressed when taking practice tests. This risk is managed by giving you three opportunities to improve your score. In between each test opportunity, you will either participate in or listen to, how students solved certain test questions in order to become better test-takers. You might feel embarrassment about your own test score results. This risk is managed by making sure that no one knows anyone else’s test results. Instead of focusing on your scores, we’ll focus on how the class did on test questions. Student names will not be shared, only the last four digits of your student ID number will be used so you can see how you did but no one else can tell how you did.

Benefits: You may benefit from participating by learning what types of test questions will be on the English language arts section of the CAHSEE as well as learning strategies to solve certain test questions. The benefits to scientific knowledge that may result are that we may learn how students think when approaching certain test questions that may impact how teachers teach test preparation in the future. We may also discover ways to help students throughout California prepare for and pass the English Language Arts section of the CAHSEE.

Discuss With Family: Please talk to your parents about this study before you decide to participate. I am sending your parents an informed consent letter and asking them to indicate whether or not they want you to participate in the study. I will need your parent(s) support in order to allow you to participate in this study. Even if your parents say that you can be in the study, you can still decide not to participate.

Questions About The Study: Please ask me any questions about the study and I will try to answer them for you. If you have questions that you think of later, please email me at moshea@sandi.net or call me at 858-496-8370 X4513. If I am not in, please leave a message and I will return your call.

SAN DIEGO STATE UNIVERSITY
INSTITUTIONAL REVIEW BOARD
Approved By: ____________________________
Expires: 2/10/2010
If you choose to participate in this study, you can withdraw at any time. If you decide not to participate now, you cannot be added once the study begins. No matter what choice you make, your grade in the class will not be affected in any way.

Please select one of the following options:

☐ I would like to participate as a participant in the CAHSEE study. I give consent for Mrs. O’Shea to analyze my test score results, my comments during class discussions, and my responses during journal reflections.

☐ I would not like to participate as a participant in the CAHSEE study. Even though I will take practice tests, listen to certain test questions during class, and write reflections in my test journal, I do not want my test data used or my discussion comments and journal entries analyzed.

Student’s Name _______________________________ Date ________________
(please print)

Student’s Signature _____________________________________________

Project Representative ________________________________ Date ________________
San Diego State University

Assent to Participate in Research as a Focus Student

Can A Constructivist Approach Be Useful To Prepare Students for Standardized Tests?
How Ninth Grade English Language Arts Students Prepared For
The California High School Exit Exam (CAHSEE).

Dear Student:

If you decide to participate in Mrs. O'Shea's research study, your parent(s) agree to let you participate, and you are an African American or Hispanic student who tested proficient, advanced, below basic, or far below basic on the 8th grade California State Standards Test in English Language Arts, you are eligible to become a focus student in the study. A total of six focus students will be selected for this study; three African American/Hispanic students who are high test-takers (those scoring proficient or advanced) and three African American/Hispanic students who are low test-takers (those scoring below basic or far below basic) on the CSTs in English.

Focus students will do everything a participating student does (take practice CAHSEE tests, participate in discussions, and write in their journals), however, these students will also participate in think alouds in which they test in a private setting (the library) and explain what they are thinking into a tape recorder about certain test questions while they test. In between tests, all focus students will be shown transcripts (typed words of what students said during think alouds) to see if they can recognize the behaviors of strong test-takers. For example, an effective test-taker might go through all of the test answer choices before making a decision, narrow down the test question to only two choices, analyze the key words in the stem of the test question, and double-check the text before making a decision.

At the end of the study, focus students will be interviewed to see which strategies (taking practice CAHSEE tests, participating in think alouds, discussing test questions, writing reflective responses) they think were the most helpful and least helpful for them to prepare for the CAHSEE. If you decide you want to participate as a focus student, please understand that you may not be selected due to the small number of students needed for this part of the study. If you are not sure whether or not you are eligible, please ask me.

Since current CAHSEE data reveals that African American and Hispanic students are the two ethnic groups most likely to fail the CAHSEE, it is important to see how those who score below basic behave during test-taking. It is also equally important to show how proficient and advanced African American and Hispanic students behave while they test in order to create a bank of behaviors that effective African American and Hispanic test-takers do to set a positive example for their peers.
Focus Student Requirements: If you are selected as a focus student (one of six African American/Hispanic students) and you agree to participate, you will be asked to complete the same tasks in the original assent letter with the following four tasks added:

- Participate (actively speak) during classroom discussions about test questions
- Record thought-processes via audiotape on certain test questions (those marked with a Princeton logo) while taking the three, 25 question Princeton Practice CAHSEE tests
- Review transcripts of you and your peers’ think aloud transcripts to see how you and your peers are responding to specific test questions
- Participate in an interview at the close of the study in which you explain which aspects of the study (taking practice CAHSEE tests, participating in think alouds, discussing test questions, writing reflective responses) were the most and least helpful.

Risks: Risks associated with the study are that some students might feel stressed when speaking into an audiotape recorder while testing. This risk is managed by giving you a private setting in which to test. In between each test opportunity, you will also be able to see how other students participating in think alouds responded to similar test questions. However, you will not know the identity of those other students since their individual transcripts will be given a number and will not contain a student’s name. Other risks might include embarrassment about your own think aloud transcripts. This risk is managed by making sure that no one knows anyone else’s transcript identity. Instead of focusing on your transcripts, we’ll focus on how the entire group of think aloud students approached the test questions. Student names will not be shared so you can see how you responded but no one else can tell how you did.

Benefits: You may benefit from participating by understanding what types of test questions will be on the English language arts section of the CAHSEE as well as learning strategies to solve certain test questions. You will also be able to see how other effective test-takers think when they approach a test question. The benefits to scientific knowledge that may result are that we may learn how students think when approaching certain test questions that may impact how teachers teach test preparation in the future. We may also discover ways to help students throughout California prepare for and pass the English Language Arts section of the CAHSEE.

Discuss With Family: Please talk to your parents about this study before you decide to participate. I am sending your parents an informed consent letter and asking them to indicate whether or not they want you to participate in the study. I will need your parent(s) support in order to allow you to participate in this study. Even if your parents say that you can be in the study, you can still decide not to participate.

Questions About The Study: Please ask me any questions about the study and I will try to answer them for you. If you have questions that you think of later, please email me at moshea@sandi.net or call me at 858-496-8370 X4513. If I am not in, please leave a message and I will return your call.
If you choose to participate as a focus student in this study, you can withdraw at any time. If you decide not to participate now, you cannot be added once the study begins. No matter what choice you make, your grade in the class will not be affected in any way.

Please select one of the following options:

☐ I would like to participate as a focus student in the CAHSEE study in which I take practice tests, discuss certain test questions during class, and write reflections in my test journal. I give consent for Mrs. O’Shea to use my think aloud transcripts for review with other students who participate in think alouds also. I also am willing to participate in an interview at the conclusion of the study.

☐ I would not like to participate as a focus student in the CAHSEE study. Even though I will take practice tests, listen to certain test questions during class, and write reflections in my test journal, I do not want to participate in think alouds or an interview at the conclusion of the study.

Student’s Name __________________________________________ Date __________________________ (please print)

Student’s Signature __________________________________________

Project Representative __________________________________________ Date __________________________
APPENDIX C

PRINCETON REVIEW PRACTICE CAHSEE TEST

FORMATS FOR THIS STUDY
## Test A Reconfiguration: 25 Questions Total

<table>
<thead>
<tr>
<th>Orig. Test Version</th>
<th>Rdg. Select #</th>
<th>? #s</th>
<th># of ?s</th>
<th>Title of Reading Selection</th>
<th>Passage Type</th>
<th>New Test Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>5</td>
<td>19-22</td>
<td>4</td>
<td>Staff Responsibilities/Bookstore Phone Operator</td>
<td>Functional Documents</td>
<td>Test A</td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>47-52</td>
<td>6</td>
<td>Writing a Short Story</td>
<td>Literary Texts</td>
<td>Test A</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>57-59</td>
<td>3</td>
<td>The Loch Ness Monster</td>
<td>Info. Passages Sci. and Nature</td>
<td>Test A</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>15-18</td>
<td>4</td>
<td>Saving The Shrinking Forest</td>
<td>Info. Passages Social Studies</td>
<td>Test A</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>18-21</td>
<td>4</td>
<td>The Beach Day</td>
<td>Literary Texts</td>
<td>Test A</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>27-30</td>
<td>4</td>
<td>Bamboo</td>
<td>Info. Passages Sci. and Nature</td>
<td>Test A</td>
</tr>
</tbody>
</table>
### Test B Reconfiguration: 25 Questions Total

<table>
<thead>
<tr>
<th>Orig. Test Version</th>
<th>Rdg. Select #</th>
<th>? #s</th>
<th># of ?s</th>
<th>Title of Reading Selection</th>
<th>Passage Type</th>
<th>New Test Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>26-28</td>
<td>3</td>
<td>On Being A Volunteer Director</td>
<td>Functional Documents</td>
<td>Test B</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>29-31</td>
<td>3</td>
<td>Reaching the Top</td>
<td>Literary Texts</td>
<td>Test B</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>53-58</td>
<td>6</td>
<td>The Incredible Northern Lights</td>
<td>Info. Passages Sci. and Nature</td>
<td>Test B</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>31-35</td>
<td>5</td>
<td>The Iowa Coast</td>
<td>Info. Passages Social Studies</td>
<td>Test B</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1-4</td>
<td>4</td>
<td>A Visit To The Doctor's Office</td>
<td>Literary Texts</td>
<td>Test B</td>
</tr>
</tbody>
</table>
Test C Reconfiguration: 25 Questions Total

<table>
<thead>
<tr>
<th>Orig. Test Version</th>
<th>Rdg. Select #</th>
<th># of ?s</th>
<th>Title of Reading Selection</th>
<th>Passage Type</th>
<th>New Test Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>6</td>
<td>23-26</td>
<td>Welcome To Our Company/Cake Director</td>
<td>Functional Documents</td>
<td>Test C</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>32-34</td>
<td>Looking Back</td>
<td>Literary Texts</td>
<td>Test C</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>36-39</td>
<td>The Pros and Cons of a Vegetarian Diet</td>
<td>Info. Passages Sci. and Nature</td>
<td>Test C</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>10-14</td>
<td>Springtime in Washington</td>
<td>Info. Passages Social Studies</td>
<td>Test C</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>41-45</td>
<td>The Big Walk</td>
<td>Literary Texts</td>
<td>Test C</td>
</tr>
</tbody>
</table>
Passages NOT Included in the A, B, C Test Reconfiguration with Omission Rationale

<table>
<thead>
<tr>
<th>Orig. Test Version</th>
<th>Rdg. Select #</th>
<th>? #s</th>
<th># of ?s</th>
<th>Title of Reading Selection</th>
<th>Passage Type</th>
<th>Reason for Omission</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1-9</td>
<td>9</td>
<td>Pride and Prejudice</td>
<td>Literary Texts</td>
<td>Too many items (&lt;7) Not enough like excerpts (&gt;3)</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>10-17</td>
<td>8</td>
<td>Audition Tips</td>
<td>Functional Documents</td>
<td>Too many items (&lt;7) Not enough like excerpts (&gt;3)</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>40-43</td>
<td>4</td>
<td>Ozymandias</td>
<td>Poetry</td>
<td>Not enough like excerpts (&gt;3)</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>44-46</td>
<td>3</td>
<td>Ms. Karpinska’s Big Change</td>
<td>Literary Texts</td>
<td>Omitted from New Test Version</td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>59-65</td>
<td>7</td>
<td>Grammar—Effective Substitutions for Underlined Parts</td>
<td>Functional Documents</td>
<td>Not tied to a reading excerpt</td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>66-72</td>
<td>7</td>
<td>Grammar—Words or Phrases that Complete Sentences</td>
<td>Functional Documents</td>
<td>Not tied to a reading excerpt</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>5-9</td>
<td>5</td>
<td>There Used to Be a Forest Here</td>
<td>Poetry</td>
<td>Not enough like excerpts (&gt;3)</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>10-14</td>
<td>5</td>
<td>Springtime in Washington</td>
<td>Info. Passages Social</td>
<td>Would not fit into the 25 question test configuration</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>49-50</td>
<td>2</td>
<td>Writing a Personal Essay</td>
<td>Functional Documents</td>
<td>Too few items (&gt;3)</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>51-52</td>
<td>2</td>
<td>Boy’s Volunteer Experience at an Animal Shelter</td>
<td>Literary Texts</td>
<td>Too few items (&gt;3)</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>60-72</td>
<td>12</td>
<td>Phytoplankton: Food for All</td>
<td>Info. Passages Sci. and Nature</td>
<td>Too many items (&lt;7) Not enough like excerpts (&gt;3)</td>
</tr>
</tbody>
</table>
APPENDIX D

INDIVIDUAL JOURNAL RESPONSE SHEET

(HYPOTHETICAL)
Post Discussion Journal Reflection Questions

After our discussion today on question 3, please:

1. Write down the question in your journal exactly as it appears.

2. Explain why you missed the question when you first took the test. Be as concrete as possible in your explanation (examples: misreading the question, not reading all of the choices, not comprehending the reading material, etc.).

If you DID NOT miss question 3, please explain, based on our discussion, how someone could have missed that test question. What misconceptions might a person have?

3. If you were to see question 3 on a test again, how might you respond differently to it based on our discussion today?

If you DID NOT miss question 3, please explain how, based on our discussion, someone might now approach the question differently if they were to see it again.
APPENDIX E

INTERVIEW QUESTIONS TO BE ASKED TO FOCUS STUDENTS AT THE COMPLETION OF THE INTERVENTION
Interview Questions to Be Asked to Focus Students at the Completion of the Intervention

The numbered areas are general topic areas and the checkmarks are potential probe areas that may or may not be explored during the interview depending on the interviewee’s responses.

1. What is your understanding of the term think aloud? What are you supposed to do?
   ✓ Explain the definition of think aloud in your own words.
   ✓ Are you supposed to say everything you think when you think aloud?
   ✓ Are you supposed to make statements extra clear for someone else to understand?
   ✓ Are you supposed to withhold some of the things you think that may not be correct?

2. In looking over your test scores for March, April, and May, what happened? Why do you think this is so?
   ✓ Did your scores go up?
   ✓ Did your scores go down?
   ✓ Did they go up and down or down and up?
   ✓ Explain external factors that may have affected your performance (physical/emotional health, test environment, time of year, other standardized tests, etc.)
   ✓ Explain internal factors that may have affected your performance (test readings were too difficult, test readings were too unfamiliar, questions were confusing, etc.)

3. In looking at just your think aloud question responses for March, April and May, what happened? Why do you think this is so? Did you try more on the think aloud questions than the regular questions? Why/why not?
   ✓ Did your scores go up?
   ✓ Did your scores go down?
   ✓ Did they go up and down or down and up?
   ✓ Explain external factors that may have affected your performance (physical/emotional health, test environment, time of year, other standardized tests, etc.)
   ✓ Explain internal factors that may have affected your performance (test readings were too difficult, test readings were too unfamiliar, questions were confusing, etc.)

4. Tell me about a typical test preparation session in which you took a practice Princeton Review test.
   ✓ What was it like reading the selections?
   ✓ How did you arrive at your answers?
   ✓ Did you go back into the text and look for the answers?
   ✓ Did you read the test questions first?
• What did you do when you could not figure out the answer?
• Did your experience taking practice Princeton Review tests change? If so, how? If not, why?

5. Tell me what it was like to participate in the Think Aloud.
• What was it like speaking into the tape recorder during the tests?
• What was it like expressing your thoughts verbally while testing?
• What was your impression when you looked back at your transcripts?
• Did your commentary change as you took subsequent tests? If so, how?

6. Tell me about a typical test preparation session in which you participated in an in-class discussion about particular test questions.
• What was it like listening to others talk about their thinking processes?
• Did you learn anything from other people’s shares? If so, what?
• Did you volunteer to the group? Why/why not?
• If so, do you think you helped others learn from your shares? If so, why?
• Did your participation in the in-class discussions change over time? If so, why?
• Did you like the fact that students’ names were left off of the print-out sheet? Why/why not?

7. Tell me about a typical test preparation session in which you wrote reflectively in your journal.
• What was it like trying to describe your thought-process in writing?
• What was it like looking back at your explanations over time?
• Did you see patterns in your thinking? If so, what were they?
• Did you benefit from seeing other sample journal entries? Why or why not?
• Did your ability to explain your thinking change over time? Why or why not?

8. When you look back on the entire intervention, what part was the most helpful for you (taking practice tests, thinking aloud, listening to or participating in in-class discussions, writing down your thought processes)? Explain why.

9. When you look back on the entire intervention, what part was the least helpful for you (taking practice tests, thinking aloud, listening to or participating in in-class discussions, writing down your thought processes)? Explain why.

10. What advice would you give younger students about preparing for standardized tests like the CAHSEE?

11. What advice would you give teachers? Can you think of other things teachers can do to make test-preparation more enjoyable?

Thank you for taking the time to meet with me!
APPENDIX F

SAMPLE MATRIX FOR ITEM ANALYSIS

(HYPOTHETICAL)
Period 2 Item Analysis of Reading Selection #1: A Visit to the Doctor’s Office

<table>
<thead>
<tr>
<th>#</th>
<th>SID#</th>
<th>Class</th>
<th>Gend.</th>
<th>Grd.</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
</tr>
</thead>
<tbody>
<tr>
<td>KEY</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>1</td>
<td>1270</td>
<td>Reg.</td>
<td>M</td>
<td>9</td>
<td>C</td>
<td>B</td>
<td>C</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>2071</td>
<td>Reg.</td>
<td>M</td>
<td>9</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>2160</td>
<td>Reg.</td>
<td>M</td>
<td>9</td>
<td>D</td>
<td>B</td>
<td>C</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>2210</td>
<td>Reg.</td>
<td>M</td>
<td>9</td>
<td>C</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>5</td>
<td>1111</td>
<td>Reg.</td>
<td>M</td>
<td>9</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>B</td>
</tr>
<tr>
<td>6</td>
<td>2153</td>
<td>Reg.</td>
<td>F</td>
<td>9</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>7</td>
<td>2589</td>
<td>Reg.</td>
<td>F</td>
<td>9</td>
<td>C</td>
<td>B</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>8</td>
<td>2453</td>
<td>Reg.</td>
<td>F</td>
<td>9</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>9</td>
<td>2209</td>
<td>Reg.</td>
<td>M</td>
<td>9</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>10</td>
<td>1492</td>
<td>Reg.</td>
<td>F</td>
<td>9</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>11</td>
<td>1095</td>
<td>Reg.</td>
<td>F</td>
<td>9</td>
<td>C</td>
<td>B</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>12</td>
<td>3326</td>
<td>Reg.</td>
<td>F</td>
<td>9</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>13</td>
<td>2589</td>
<td>Reg.</td>
<td>M</td>
<td>9</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>B</td>
</tr>
<tr>
<td>14</td>
<td>2177</td>
<td>Reg.</td>
<td>M</td>
<td>9</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>B</td>
</tr>
<tr>
<td>15</td>
<td>2389</td>
<td>Reg.</td>
<td>F</td>
<td>9</td>
<td>C</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>16</td>
<td>2047</td>
<td>Reg.</td>
<td>F</td>
<td>9</td>
<td>C</td>
<td>B</td>
<td>C</td>
<td>B</td>
</tr>
<tr>
<td>17</td>
<td>2318</td>
<td>Reg.</td>
<td>M</td>
<td>9</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>B</td>
</tr>
<tr>
<td>18</td>
<td>1304</td>
<td>Reg.</td>
<td>F</td>
<td>9</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>B</td>
</tr>
<tr>
<td>19</td>
<td>2316</td>
<td>Reg.</td>
<td>M</td>
<td>9</td>
<td>C</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>20</td>
<td>1627</td>
<td>Reg.</td>
<td>F</td>
<td>9</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>21</td>
<td>2296</td>
<td>Reg.</td>
<td>M</td>
<td>9</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>B</td>
</tr>
<tr>
<td>22</td>
<td>2149</td>
<td>Reg.</td>
<td>M</td>
<td>9</td>
<td>C</td>
<td>B</td>
<td>B</td>
<td>D</td>
</tr>
<tr>
<td>23</td>
<td>2182</td>
<td>Reg.</td>
<td>M</td>
<td>9</td>
<td>A</td>
<td>C</td>
<td>B</td>
<td>D</td>
</tr>
<tr>
<td>24</td>
<td>1080</td>
<td>Reg.</td>
<td>M</td>
<td>9</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
</tbody>
</table>

| ITEM ANALYSIS | 22 | 14 | 5 | 20 |
| Percentage Correct | 91.7 | 58.3 | 20.8 | 83.3 |
APPENDIX G

GROUP DISCUSSION TRANSCRIPT SAMPLE

QUESTIONS (HYPOTHETICAL)
1. After adding up the class scores, which test question did we score highest one? Lowest on?

2. Which three would you, as a class, like to review? Let’s vote by secret ballot.

3. Okay number 3. Only 5 out of 24 of us got number 3 correct. Would anybody who got number 3 correct like to explain his/her processes? If students do not respond initially, then the teacher will model by highlighting the process of examining the stem of the test question, eliminating incorrect answers first, and then double checking the selected answer against the original test question.

4. Once a student is willing to volunteer, the teacher researcher will ask guiding questions, such as “What is the test question asking, in your own words?” “Which choices are incorrect and how do you know?” “If you are not sure about two answers, what will you do to make a decision?”