Getting Inside the Mind of a Student Preparing for the SAT

Alex Vigderman

University of Pennsylvania

December 31, 2010

Abstract

The effectiveness of SAT preparation programs has been well-studied, but the preparation process has not been examined in past research from the perspective of the student. The nature of students' subjective experience while preparing for the SAT was of greatest interest, as well as how committing different kinds of errors on practice questions relates to actual test outcomes. As an exploratory study, a sample of students ($N = 17$) beginning a one-on-one tutoring program completed diagnostic SAT practice math questions during which they were asked to think aloud; errors were categorized based on information gleaned from the participants' thought processes. Errors from misunderstanding the wording or dominant mathematical concept of a question were most common; however, only conceptual error frequency predicted both initial and final practice SAT scores. Participants reported low enjoyment and interest and high importance, concentration, and challenge level in SAT preparation activities as compared to a selection of common activities. The degree to which participants felt challenged by SAT preparation inversely predicted performance on their final practice SAT math section when controlling for their initial practice math section score. This suggests that prior competence in the form of perceived ease of test preparation predicts improvement on the SAT.

Getting Inside the Mind of a Student Preparing for the SAT

The viability of the SAT exam as a measure of student aptitude as well as its susceptibility to coaching has been a subject of recent debate. Much evidence indicates that scores can be significantly increased through coaching, and as a result the focus has turned to designing the most effective preparation designs (see Joseph, 2004). In order to improve preparation curricula, it is important to understand the dispositions and subjective experiences of students who show the most success with SAT preparation. Student performance during test preparation is likely to be affected by cognitive and emotional predispositions toward the preparation process. For example, students who do not feel that they can improve are less likely to do so (Multon, Brown, & Lent, 1991; Pajares & Kranzler, 1995). Recent research of "flow theory" (Shernoff, Csikszentmihalyi, Schneider, & Shernoff, 2003) and deliberate practice (Ericsson, 2003) provides a basis for the analysis of optimal student performance, but neither of these theories has been applied to standardized test preparation. Analysis of the thought processes of students while they complete SAT problems through protocol analysis (Ericsson, 2006) can highlight aspects of questions that coaches must address to improve student performance.

SAT Coaching Efficacy

While reports commissioned by the College Board claim that the effects of SAT coaching are overstated (e.g., Powers & Rock, 1998), the benefits of coaching are nonetheless undeniable. Powers and Rock (1998) reported mean gains of 20 points on the Math section for students enrolled in two of the most well-known coaching companies (Princeton Review and Kaplan) compared to students who were not enrolled with those companies. In a similar observational study of students enrolled in Princeton Review and Kaplan programs, Smyth (1989) reported gains of 39 and 26 points on the SAT Math section when compared to non-coached students. However, this study is subject to criticism because the reported increase for students that did not engage in coaching (38 points on the Math section) was much higher than the expected gain from simply retaking the exam reported by the College Board (College Board, 2010).

In a more tightly controlled pre-test/post-test design with over 30 hours of preparation for the Mathematics section of the PSAT exam, Reynolds, Oberman, and Perlman (1988) reported that coached students scored 41 points better than non-coached students on the Math section of the PSAT post-test, and this effect increased to 68 points for groups that used actual past PSATs as practice. Matthew Joseph, who was consulted for the design of the current study, implemented one-on-one tutoring sessions including specialized worksheets and past SAT exams for practice and reported average gains of about 40 points on the Math section for coached students compared to non-coached students in a socioeconomically diverse sample (Joseph, 2004). Despite limitations emerging from nonresponse on homework packets, this approach included protocols for tutoring that were explicitly defined and controlled, which cannot be said for the observational studies above, and the score improvement effects were greater than those for less tightly controlled studies.

Perceptions of Competency

Given that coaching is effective, it is important to determine those aspects of student personality and the tutoring experience that make coaching most effective for recipients. One important trait is a student's perception of how difficult the subject is, which can arise from his/her self-efficacy expectations (Bandura, 1977) and internal locus of control (Ajzen, 2002). These constructs are contained within a broader range of dispositions known as core selfevaluations, which also include self-esteem and dispositional optimism (Judge, Locke, & Durham, 1997).

Self-efficacy. Bandura (1977) posits that the strength of a person's efficacy expectations (i.e., expectations about one's ability to execute a behavior that will bring about some outcome) determines how much effort and persistence he will expend to achieve some goal, and thus is important to predicting performance outcomes. These efficacy expectations are domain-specific and are formed through a combination of previous performance accomplishments, vicarious experience of success, verbal persuasion, and emotional arousal during stressful events (Bandura, 1977). In school, students receive efficacy input from their own performance, their peers' performance, instructor feedback, and anxiety responses, which impact how effective they perceive themselves to be in studying or preparing for a test such as the SAT. When students feel that a topic is too difficult for them, their self-efficacy in terms of being able to learn that topic decreases, and they will be less likely to put in the requisite effort to actually learn it.

Greater self-efficacy expectations have been shown to produce stronger academic performance, especially in the field of mathematics. Pajares and Kranzler (1995) showed through path analysis that self-efficacy in mathematics predicted performance on a standardized set of math problems over and above general mental ability (as measured by Raven's Advanced Progressive Matrices), math anxiety, and high school math level. A meta-analysis of self-efficacy research concluded that self-efficacy beliefs accounted for about 14% of the variance in students' academic performance (Multon et al., 1991). This analysis found that effect sizes for self-efficacy's effect on performance were increased for low-achieving groups and older students, both of which are more likely to benefit from a self-efficacy intervention than high-achieving or elementary school students (Multon et al., 1991). Students who enroll in SAT preparation are not likely to be top-tier achievers, so they should show a large effect of self-efficacy on SAT performance.

Internal locus of control. Internal locus of control is the degree to which a person believes that outcomes in her life directly result from decisions she makes (Ajzen, 2002). This construct mirrors Bandura's conceptualization of outcome expectations, "a person's estimate that a given behavior will lead to certain outcomes" (Bandura, 1977, p. 193). Burger (1989) points out a significant aspect of internal locus of control in that this control need not actually exist, but it is the perception of control that creates a particular response. He also suggests that this perceived control can have negative effects when one becomes concerned about self-presentation, which is especially present in cases of self-handicapping, or the likelihood of a negative outcome regardless of control (Burger, 1989). Internal locus of control has been connected to varied external benefits such as job satisfaction and performance (Judge, Locke, & Durham, 1997) and coping skills (Jackson & Coursey, 1988).

Students' perceptions of how much control they have over their academic performance may result from the degree to which their grades are predictable (Burger, 1989). Perceived control in this sense would relate to the degree to which a student's grades correlate with the amount of effort expended on work for that assignment or test. Ward and Jenkins (1965), however, showed that people perceive their level of control to be greatest when the probability of success is greatest because we wish to believe that positive outcomes are brought about as a result of our actions. From the Ward/Jenkins research, it can be concluded that students will show increased perceived internal locus of control when they do well in a subject, regardless of the amount of effort put into attaining success. Therefore, internal locus of control is subject to the same prerequisite conditions as self-efficacy (namely, previous success), and a combination of the two constructs into a broad umbrella of competence is appropriate. This competence may be a result of relevant coursework being easy or the student having relative skill, but the impact of this disposition on academic motivation is that the student feels that the subject comes more easily to him.

The Subjective Experience of Optimal Performance

The flow concept. In the last twenty years or so, the field of positive psychology has blossomed and created interest in the psychology of positive experience. One of the most important areas of research in positive psychology is "flow," proposed by Mihalyi Csikszentmihalyi (Csikszentmihalyi, 1990). Csikszentmihalyi conceptualizes flow as a state of intense concentration and absorption in an activity that is intrinsically enjoyable and can only be produced when a delicate balance of high individual skill and task challenge is struck (Csikszentmihalyi, 1997). In operationalizing flow, researchers have identified the following states as important hallmarks of a flow experience: concentration, enjoyment, positive affect, feelings of strength, desire to be doing the activity, high self-esteem, and a sense that the activity is important to future goals (Hektner & Asakawa, 2000). Flow has been associated with a higher quality of experience in sports, chess, dancing, music, and many other activities which people perform for intrinsic benefit (Csikszentmihalyi & Csikszentmihalyi, 1988).

Research on flow's place in academic settings has focused primarily on three dimensions of flow: concentration, interest, and enjoyment (Shernoff et al., 2003). Shernoff et al. propose that this subset of the hallmarks of flow represents student engagement in an academic setting. These psychological states are captured using the Experience Sampling Method (ESM), in which participants are randomly "beeped" several times throughout the course of a day or week and are asked to fill out a questionnaire about their physical and psychological state at the time they were beeped (Csikszentmihalyi & Larson, 1987). In their 2003 study, Shernoff et al. used ESM data from the Sloan Study of Youth and Social Development (see Csikszentmihalyi & Schneider, 2000) to assess engagement, skill, and challenge levels for the in-school activities that were reported in the study. Student engagement was at its highest when students were involved in an activity that balanced high skill with high challenge, which is referred to as a flow activity (Shernoff et al., 2003, p. 167). Most of the activities which were categorized as flow-like involved individual or group work which required more interactive learning compared to lectures, videos, or exams.

Deliberate practice. Beyond studies based in flow theory, there is evidence that increased student engagement promotes academic achievement (e.g., Singh, Granville, & Dika, 2002; Finn & Rock, 1997). In schoolwork, unfortunately, students do not achieve as much enjoyment as would be expected from an optimally engaging experience (Shernoff et al., 2003). As such, there is room for other models of optimal academic performance. One such model is Ericsson's method of deliberate practice, in which expert performance is attained not through exceptional talent but through time-consuming, structured attempts to address weaknesses (Ericsson, Krampe, & Tesche-Römer, 1993). Ericsson's ideas have been supported by research with expert medical professionals (Ericsson, 2004), pianists (Krampe & Ericsson, 1991), and chess players (Charness, Tuffiash,, Krampe, Reingold, & Vasyukova, 2005).

In the context of academics in particular, Sternberg (1998) argues that youths that become expert students have developed various important cognitive mechanisms (e.g., have cultivated metacognitive skills, created large and organized schemas of declarative knowledge, and found ways to efficiently draw on these stores of knowledge when appropriate). By

Ericsson's criteria, a student could become an expert by spending a great deal of time continually solving problems that represent critical junctures in his learning curve such that his performance goes beyond the arrested development stage at which most people would peak (Ericsson, 2003). Put differently, the student can work to reduce the difference between his skill level and the challenge level of the task, and once the two are equal he can seek a more difficult goal to work towards. Deliberate practice thus interacts with flow theory, because deficits in skill can be remedied through deliberate practice such that skill can be balanced with challenge and a flow state can be attained. Ericsson, who was consulted for the present study, feels that this collaborative model provides evidence that while some aspects of academic experience may be flow-like (e.g., exams, group activities), the process of studying and completing homework is more akin to deliberate practice.

If studying and completing homework (e.g., for the SAT) is a form of deliberate practice, Ericsson's research predicts that the experience of studying for the SAT would involve a great deal of concentration, interest, and task difficulty, but would differ from a flow-like conception in that the student's enjoyment and skill would be low. By Csikszentmihalyi's model, a low-skill/high-challenge activity of this type would be likely to produce anxiety instead of flow, and this anxiety would in turn act as a motivating factor for increasing one's skill (i.e., engaging in deliberate practice) or softening the challenge (Csikszentmihalyi, 1997). Therefore, we can ascertain the true nature of SAT studying by inquiring of participants the aspects of the ESM form that should serve to connect (e.g., concentration, interest, challenge) or distinguish (skill, enjoyment) deliberate practice and flow.

Protocol analysis. The most difficult task in assessing a person's subjective experience while performing a task is reading the mind of the performer as the task is being conducted, as opposed to immediately afterward (which is the case for ESM research). Ericsson and Simon (1984) developed protocol analysis for this purpose, which is a method of eliciting verbal reports of thoughts as they occur without receiving reactive or descriptive statements about the courses of action taken. As Ericsson (2006) puts it:

> The central assumption of protocol analysis is that it is possible to instruct subjects to verbalize their thoughts in a manner that does not alter the sequence and content of thoughts mediating the completion of a task and therefore should reflect immediately available information during thinking. (p. 227)

Thus, if instruction is given appropriately, the performer will be speaking with such a short delay after actually having the thought that there will not be sufficient time for him/her to attempt to interpret or describe the reasons for certain thoughts. Ericsson and Simon (1993) were unable to find evidence that the thought sequence, and thus performance accuracy, was affected by the addition of think-aloud protocols to tasks.

Protocol analysis has been applied in several of the expert domains that Ericsson researches (see Ericsson, 2004; Krampe & Ericsson, 1991; Charness et al., 2005). However, its applications to education are potentially significant. Having accurate representations of the thoughts mediating math problem solving in students preparing for the SAT can allow tutors to assess the particular aspects of problem-solving or the questions themselves that students are failing to properly encode. While there have been attempts to code the types of errors that students make on math problems (e.g., Elbrink, 2008), the insights that protocol analysis can provide into students' thought processes are invaluable resources for curriculum development in a one-on-one tutoring environment. Combining verbal reports with information gleaned from notes on scratch paper can allow instructors to categorize errors and choose a course of instruction that is appropriate for students who commit those types of errors. An important consideration when instructing children to engage in protocol analysis, though, is to ensure that no self-explanations are generated, as these descriptions often alter performance (Ericsson, 2006). While performance improvement is the desired outcome of SAT tutoring in the long run, accurate representation of the thought processes mediating incorrect answers is most important for the assessment of a student's failings.

Hypotheses to be Tested

The current study sought to explore the nature of students' subjective experience while preparing for the SAT exam, in particular their opinions about the tutoring process and their thought processes while solving math problems. Additionally, competence as measured by perceived ease of preparatory work was to be tested as a predictor of final practice SAT exam score. My first hypothesis was that participants' perceptions of the SAT preparation experience would involve intense concentration, importance, and difficulty, but would also involve low levels of skill, interest, and enjoyment as predicted by deliberate practice theory (Ericsson et al., 1993). The second aspect of the study was more exploratory; protocol analysis was to be used to ascertain the type and frequency of errors made by students enrolled in SAT preparation and determine if the frequency of certain types of errors was more predictive of SAT success and improvement than others. The last hypothesis to be tested was that perceived level of difficulty, which results from previous academic success, would predict a student's score on his/her final SAT practice test math section, controlling for initial practice test math score. The hypothesized mediator of this relationship was improvement on a series of diagnostic tests presented during the first month of tutoring, as shown in Figure 1.

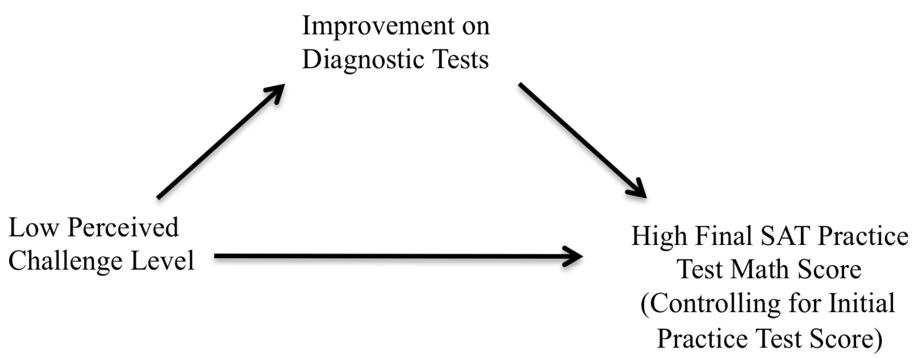

Figure 1. Visual Representation of Proposed Mediation Model.

Method

Participants

Seventeen suburban Philadelphia high school students (10 male, 7 female) completed the study. One additional student consented but chose during the process not to participate. Participants were recruited after an informational meeting with their tutor at a test preparation center in suburban Pennsylvania. Ages ranged from 15 to 17 with a mean of 16.4 years (SD = 0.42) at the time of consent. Participants were paid $50 for participating and $18 as reimbursement for their itemized SAT score report.

Procedure

Participants completed a personality questionnaire (the baseline questionnaire; see Appendix A) immediately after recruitment in their first week of tutoring. As participants completed their baseline questionnaire, parents were asked to complete a similar questionnaire about their child (see Appendix B). After the participants' second tutoring session, they completed a questionnaire (the anchor questionnaire) about their attitudes and mental states while engaging in a selection of activities (see Appendix C). After the participants' second and third sessions, they completed one set of fifteen diagnostic questions about Ratios per session; after their fourth and fifth sessions, participants completed one set of fifteen diagnostic questions about Averages per session. For each of these diagnostic tests, the participants wore a headset and were asked to think aloud as they solved problems. After the participants had completed their program with the test preparation center, the score for each participant's first and last practice SAT (administered as part of the center's normal curriculum) were recorded as additional baseline and outcome variables.

Timeline. Tutors were given a projected timeline, which was created by the experimenters, for the participants' progress through the different subjects being taught. This timeline stressed in particular when the participant should work on Ratio and Average problems. The experimenters were told prior to the start of the study that these subjects would be taught universally among the participants.

While many of the participants had weekly tutoring sessions, some participants had irregular schedules as a result of being involved in other activities. The present study took place over the first five sessions of the participants' fourteen sessions at the test preparation center, by which point students have usually taken the SAT. The diagnostic tests for Ratios and Averages were administered in a pre-test/post-test format such that tutors would review strategies for the subject in question in the time between the pre-test and post-test for that subject. The schedule for each participant was based on the order of his/her sessions, and adhered to the following format:

Session 1.

Before session. The experimenters spoke with the prospective participant and parents prior to their meeting with the tutor to briefly explain the requirements and goals of the study.

During session. The prospective participant and parents hold informational meeting with the tutor.

After session. Consenting participants and parents filled out consent forms and baseline questionnaires.

Session 2.

Before session. During the week between Session 1 and Session 2, the participant completed an initial SAT practice test as a standardized baseline score.

During session. The tutor worked with the participant on a subject other than Ratios and Averages, and assigned as homework a packet of Ratio problems.

After session. An experimenter met with the participant and read from a script that described the appropriate protocols for thinking aloud concurrently with completion of diagnostic test problems. The experimenter gave participants non-academic practice exercises to ensure that they properly understood the process of thinking aloud (see Appendix D). Following, the participant completed the Ratio diagnostic pre-test. After this examination was complete, the participant completed the anchoring questionnaire.

Session 3.

Before session. No intervention.

During session. The tutor reviewed Ratio homework packet and discussed strategies for solving Ratio problems. The tutor assigned homework in a subject other than Ratios and Averages.

After session. The participant completed the Ratio diagnostic post-test.

Session 4.

Before session. No intervention.

During session. The tutor worked with participant on a subject other than Ratios and Averages, and assigned as homework a packet of Average problems.

After session. The participant completed the Average diagnostic pre-test.

Session 5.

Before session. No intervention.

During session. The tutor reviewed Average homework packet and discussed strategies for solving Average problems.

After session. The participant completed the Average diagnostic post-test.

Participants completed a final SAT practice test in the week prior to their last session (approximately their 14^{th} session) which was used as an outcome variable. Upon completion of the actual SAT and receipt of the itemized score report, participants were given their monetary compensation in exchange for a copy of the report.

Diagnostic tests and audio coding. For all computer-based diagnostic tests, participants sat in a room in the test preparation center with an experimenter and were given as much time as they needed to complete the fifteen-question tests. A graphing calculator and scrap paper were provided. While completing the practice problems, participants were required to wear a headset and were prompted to speak aloud each of their thoughts about the problem as they occurred to them. The experimenter remained in the room to ensure that the subject was reporting his/her thoughts as much as possible, as well as to clarify any questions about the format of the examination. If a participant asked questions about the content of the questions, no assistance was given.

Using a USB headset, the diagnostic test application recorded the subject's speech, answers, and time taken for each question, which the experimenters were able to download. The experimenters compared each participant's responses to an answer key, and recorded the number of questions each participant answered correctly on each test. For those questions that participants answered incorrectly, two experimenters listened to the audio recordings of the participants' think-aloud protocols to assess the type of error that was made for each incorrect answer given.

Measures

Anchor questionnaire. The anchor questionnaire was designed to assess the nature of participants' subjective experience when engaging in SAT preparation activities as compared to other common activities. Each of the eight pages of the questionnaire contained items about a particular aspect of subjective experience assessed on the Experience Sampling Method form (Csikszentmihalyi and Larson, 1987). The aspects of subjective experience used and their page headers are listed in Table 1.

Table 1. *Anchor questionnaire categories and page headers.*

Subjective Experience	Page Header
Enjoyment	Please rate your enjoyment of the following activities in terms of their ability to elicit in-the-moment enjoyment.
Importance to Future Goals	Please rate the importance of the following activities with regards to your future goals.
Concentration	Please rate how well you typically concentrate when performing the following activities.
Challenge	Please rate how challenging the following activities are.
Interest	Please rate how much you wish you were doing something else while doing the following activities. (reverse coded)
Control	Please rate how in control you feel when doing the following activities.
Importance to SAT Improvement	Please rate the importance of the following activities with regards to improving your SAT score.
Skill	Please rate how skilled you feel when completing the following activities.

On each page, three activities associated with SAT preparation (e.g., "Meeting with my SAT coach") were included with ten common activities for high school students to experience (e.g., "Watching television", "Waiting in a long line", "Taking a test in my best subject"). The SAT activities were chosen based on their prevalence in the test preparation center's curriculum, and the miscellaneous activities were selected evenly from representative activities of Hektner's and Asakawa's four categories of subjective experience (Hektner & Asakawa, 2000). Participants rate the degree to which they have the experience described at the top of the page when engaging in each of the activities on a 9-point scale, where 1 corresponds to not having the subjective experience at all and 9 corresponds to having a complete subjective experience of that type. Observed internal reliabilities for these measures are included in Table 2.

Diagnostic tests. The test preparation center provided all SAT preparation homework packets and tutoring that would normally be provided with their services. In addition, four 15-question practice tests about Ratios and Averages (Ratios A and B, Averages A and B) of approximately similar difficulty were compiled by staff at the test preparation center before use by the experimenters (see Appendices E-H). The order of the Ratio and Average tests that each participant was to take was pre-determined based on the order in which participants were recruited for the study, as the experimenters felt that this method of counterbalancing would be more effective than simple random selection for such a small sample. Of the 60 total questions on these tests, 25 were adapted from questions on past Graduate Management Admission Tests, 20 were adapted from questions on past SAT exams, and 15 were invented by the test preparation center to test general mastery of the subject in question. The difficulty of each question was rated to consensus on a scale of 1 to 5 by four experimenters, and the total difficulty rating for each test was either 49 or 50, indicating equivalent difficulty between tests of the same subject. These tests were compiled into an online computer application that was created for the current study.

Audio coding. Experimenters collaborated with tutors at the test preparation center to create a coding scheme that would categorize errors into six possible categories: misreading or misunderstanding the nature of the question, making a computational mistake, failing to properly understand the key concept being assessed on the test, skipping the question entirely, errors for which the experimenters could understand what the participant was saying, but could not discern a coherent thought process; and finally, errors that could not be categorized as a result of the participant speaking unintelligibly, too quietly, or not at all. Each incorrect response was coded to consensus by two experimenters (disagreements occurred on roughly one-fifth of errors) and each coder wrote down brief notes describing the error in more detail.

Results

The significance level for statistical tests was one-tailed $p < .10$ because of the exploratory nature of the research and directionality of my hypotheses.

The Subjective Experience of SAT Preparation

Students find SAT preparation to be challenging and aversive but important as compared to a selection of common activities. Table 2 shows the mean response for each subjective experience category for the aggregated SAT preparation activities as compared to the mean in other activities on the anchor questionnaire. Participants felt more challenged, experienced more concentration, and reported greater importance to both SAT performance and future goal achievement for SAT preparation activities as compared to other activities. Participants also reported lower enjoyment and interest for SAT preparation activities.

Coding the Types of Errors Students Make

Participants answered, on average, about four questions out of 15 incorrectly on diagnostic pre-tests and about three questions out of 15 incorrectly on diagnostic post-tests. This reduction in errors was significant within subjects ($t(16) = 3.20$, $p < .01$), indicating that participants improved from pre-test to post-test. Table 3 shows the distribution of error types among these errors. The most frequent errors were misread and conceptual errors; these represented 31% and 27% of all errors, respectively. Participants who skipped more questions on diagnostic tests showed greater improvement on those tests ($r = .57$, $p < .05$), while participants who made more miscellaneous errors showed less improvement, $r = -.62$, $p < .01$.

Subjective Experience and Error Type as Predictors of SAT success

Participants who reported greater control ($r = .49$, $p < .05$) or less difficulty ($r = -.45$, $p < .10$) for SAT preparation activities scored higher on their final SAT practice test math section. No other anchor questionnaire items were predictive of final practice test score (see Table 4). Those who made fewer conceptual errors on diagnostic tests reported more enjoyment ($r = -.60$, $p <.05$) and control ($r = -.69$, $p < .01$) during SAT preparation activities. As shown in Table 5, participants who committed fewer conceptual errors on diagnostic tests scored higher on both their initial and final SAT practice tests. However, neither students who made fewer conceptual errors ($\beta = -.28$, *ns*) nor total errors ($\beta = -.36$, *ns*) showed higher final SAT practice test scores when controlling for first SAT practice test score. Despite the discouraging result, I suspect that the lack of significance may be a result of having insufficient statistical power.

Improvement on Diagnostic Tests as a Mediator for Challenge Level and SAT Score

My initial hypothesis about SAT improvement was that participants who reported higher perceived challenge for SAT preparation activities would have lower final SAT practice test math scores after controlling for initial scores, and this effect would be mediated by improvement on the diagnostic tests. As it turned out, students who perceived SAT preparation activities as less challenging and felt more control during those activities scored higher on their final practice tests. Since no other anchor questionnaire variables were predictive of final practice test score, only challenge and control level were considered. However, further inspection showed that students who reported high control on SAT activities did not show a higher final SAT practice test score when controlling for initial practice test score ($\beta = .02$, *ns*), and so only challenge level was considered for the mediation model.

In order for a variable to be considered a mediator, three criteria must be met: The independent variable must predict the mediator, the independent variable must predict the dependent variable, and the mediator must predict the dependent variable when the independent variable is held constant. Baron and Kenny (1986) have outlined procedures for confirming these relationships and then determining whether the relationship between the independent and dependent variables is significantly diminished when controlling for the mediator. In the present study, the predicted mediation was that participants who felt less challenged by SAT preparation activities (and thus would be more confident in their abilities) would achieve higher final practice test scores when controlling for initial practice test scores, with improvement on the diagnostic tests as a mediator. When appropriate, part correlations will be reported, which describe a relationship between predictors and outcome, controlling for the effects of other variables in the equation.

Improvement on diagnostic tests did not mediate the relationship between challenge and final practice test math score when controlling for first practice test math score. Regressing improvement on diagnostic tests on challenge level showed a non-significant relationship ($\beta = -.05$, part $r = .05$, *ns*). In a hierarchical multiple regression predicting final practice test math score while controlling for first practice test math score, challenge level was entered in Step 1 and found to be significant ($\beta = -.36$, $p < .05$), but when improvement on diagnostic tests was added in Step 2, the β for challenge level did not fall and continued to be significant ($\beta = -.35$, part $r = -.35$, $p < .05$). Refer to Table 6 for the results of this hierarchical regression and Figure 2 for a visual representation of the mediation model with associated Beta coefficients.

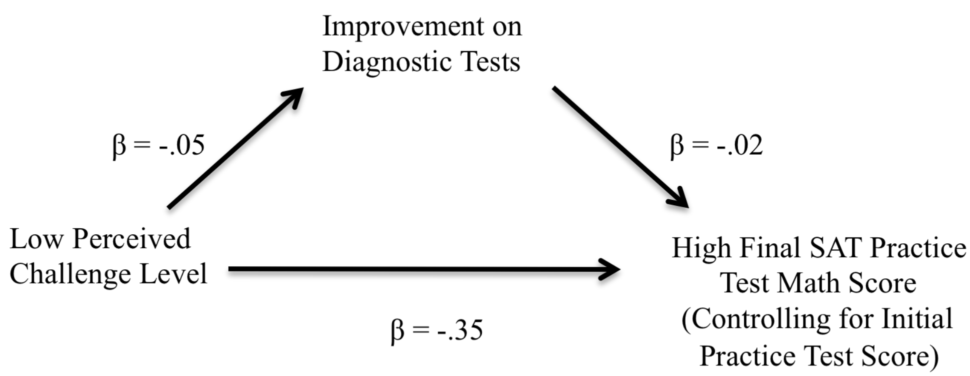

Figure 2. Visual Representation of Mediation Model with Corresponding Beta Coefficients.

Discussion

The fact that participants rated SAT preparation as involving significantly more challenge, importance to future goals and SAT improvement, and concentration than a selection of common activities supports my first hypothesis. Also supporting this hypothesis is the fact that participants reported significantly less skill, interest, and enjoyment for SAT preparation as compared to other activities. The second purpose of the study, the exploratory look at error type composition on diagnostic tests, revealed that a large proportion of errors were related to faulty conceptual knowledge, but a similar proportion of errors were a result of misunderstanding or misreading the question, which is a potentially under-emphasized type of mistake. Additionally, students who skipped more questions and committed fewer errors on these diagnostic tests that were categorized as miscellaneous (which are described in more depth below) improved their

SAT score more from the beginning to the end of their tutoring curriculum. The final hypothesis was not supported by the data; the effect of low perceived challenge level on SAT improvement was not mediated by improvement on the diagnostic tests.

Subjective Experience of SAT Preparation

The results from the anchor questionnaire items indicate that SAT preparation is more likely to be a form of deliberate practice than a flow-like experience. SAT preparation appears, like deliberate practice, to be an effortful, challenging, and focused endeavor for which the individual has to increase her skill to achieve what she wishes to accomplish. Naturally, there may be individual aspects of SAT preparation that are more or less similar to flow which are not ascertained in the current study, since Shernoff et al. report that flow-like engagement occurs roughly 15% of the time in school settings (Shernoff et al., 2003). Additionally, individual differences in autotelic personality, the degree to which individuals engage in activities for their own sake (Csikszentmihalyi, 1975), could produce much of the variability in the type of experience that SAT preparation is for students.

Participants who reported greater control for SAT preparation activities showed higher initial and final practice SAT scores, which indicates that previous mastery or experience with the SAT was a factor for improved performance. Studies on test-wiseness and familiarization paradigms (Messick, 1980; Powers & Alderman, 1983) have showed that exposing students to the SAT prior to taking the exam improves confidence and promotes advantageous SAT-specific test-taking behaviors (e.g., knowing when to guess or omit a question) without necessarily providing any coaching. These types of results are corroborated by the presence of test-retest score increases (College Board, 2010), which likely come from increased level of awareness of the types of questions and pace of the exam and not an increase in relevant content knowledge.

Perceived control during SAT preparation activities did not predict final SAT practice test score when controlling for initial score, which indicates that the degree of control was a pre-existing disposition that did not impact and was not impacted by the coaching process.

Error Types on Diagnostic Tests

Of the approximately 15 errors per participant that were made on the four diagnostic tests combined, about eight of them were categorized as a conceptual error or a misread error. The remaining errors were about evenly split among computational errors, skipped questions, and miscellaneous errors. Since most of SAT tutoring is concerned with teaching relevant conceptual content (Joseph, 2004), it is important to note that conceptual errors represent a large proportion of the mistakes that students make on these types of questions. One intriguing result was that female participants committed significantly fewer misread errors ($r = -.55$, $p < .01$); however, they did not show improved performance on the diagnostic tests on the whole or their final SAT score, and thus this difference does not appear to have practical implications.

The proportion of skipped questions was slightly inflated because some skipped questions were actually a result of the participant unknowingly clicking “Next” on the online diagnostic test form, thus not having any relation to the participant’s aptitude for that question. Additionally, further inspection of the miscellaneous category revealed that most of these errors could be categorized as either guesses, faulty knowledge of a mathematical concept that was not directly assessed in that diagnostic, or failures to properly keep track of one’s intermediate steps during a problem. While the latter two are specific kinds of errors that can be identified and worked on, the former could result from any of a host of errors and is thus difficult to eliminate. However, each of those sub-categories represents a small proportion of errors, and thus allocating a good deal of resources towards eliminating these errors may be an inefficient way to tutor.

Diagnostic Test Improvement as a Mediator for Perceived Challenge and SAT Improvement

There was no indication of a mediator relationship between low perceived challenge of SAT preparation activities, diagnostic test improvement, and final SAT practice test math score, when controlling for initial practice test math score. This result may be impacted by the fact that the diagnostic test improvement variable represented a difference between two scores but had no indicator of where the scores lay on an absolute scale. Participants who scored higher on pre-tests had less room for improvement and thus experienced a ceiling effect with their post-test scores. Some participants who showed low improvement on diagnostic tests would be more likely to report having a low perceived challenge level for SAT activities, and this would reduce the correlation between challenge level and diagnostic test improvement. Additionally, the diagnostic tests represented only a small range of mathematical concepts that are tested on the SAT, and thus performance on tests related to Ratios and Averages may not have much bearing on performance on a full-scale practice test.

Limitations

The most prominent limitation of the present study is the sample itself, because a sample size of 17 reduces the likelihood of achieving statistically significant results for practically significant effects. For example, the correlation between perceived challenge level and perceived level of control for SAT activities is -.40 (see Table 4) but is not significant even at the 90% significance level. Additionally, the sample was taken from a test preparation center in an affluent suburb, and thus the range of demographics and school resources, and abilities is restricted. There may also be a restriction of range resulting from the participants being selected from a group that was already going to enroll in SAT preparation, which may have placed a floor on variables of intrinsic interest and enjoyment.

The study design itself may have created limitations as well, because practice with homework was not strictly monitored and subjective experience with regards to SAT preparation activities was not assessed at the end of the study for comparison. Proper completion of relevant homework is an issue with tutoring in general (e.g., Joseph, 2004), but in a pre-test/post-test design with specific topics it is critical that participants spend time in between tests learning the specific material to be tested. The subjects assessed in the diagnostic tests represented only a small fraction of the types of questions that are found on the SAT, and including other subjects may have improved the generalizability and predictive value of the results. Additionally, the anchor questionnaire was not presented again at the end of the study, which would have allowed me to make within-subjects comparisons of participants' subjective experience and opinions. This information could have shown if changes in disposition with regards to the SAT predicted changes in scores.

There were a few aspects of the original study design that were excluded either from the design due to logistical issues or from this paper due to insignificant results. Weekly homework assignments included an inserted page that asked about each of the subjective experience categories included on the anchor questionnaire with regards to that particular assignment. This data, collected over a series of assignments, could have provided in-the-moment time-series data about participants' subjective experience that could not have been assessed with the anchor questionnaires. However, the actual homework assignments were not reliably returned to the tutors and tutors often gave participants packets that did not include the ESM insert, and thus there was too much missing data for the information to be useful.

The baseline questionnaire, which included measures of the Big Five personality dimensions (John & Srivastava, 1999), grit (Duckworth, Peterson, Matthews, & Kelly; 2007), and self-control (Tangney, Baumeister, & Boone; 2004), did not show significant relationships with error type frequency or SAT scores, despite these characteristics having been shown to predict academic success in the past (e.g., Duckworth et al., 2007; Noftle & Robins, 2007). Items on the questionnaire concerning competitiveness and motivation did not show relationships with the dependent variables in question either. As a result, the questionnaire was eliminated from analysis.

Further Directions and Implications

Due to the small sample size and other logistical limitations, the current study should be treated as a pilot for further research. For example, the assessment of subjective experience of students as they engage in learning activities can provide suggestions about how best to engage and motivate students (e.g., Shernoff et al., 2003). These results, combined with the data from the Sloan Study (Shernoff et al., 2003), can establish the types of experiences that students have while engaging in different kinds of academic activities (e.g., homework, lectures, group activities, exams). A taxonomy of the types of experiences that students have in different academic settings can help integrate the deliberate practice and flow theories.

A great deal of improvement in tutoring curricula can be made with the application of protocol analysis to academic work. Based on the categories of errors used here and the subcategories that emerged as the study developed, one could create a more refined set of error types and design a coaching strategy to reduce each type of error in a more systematic and informed fashion. The fact that misread errors were just as common as conceptual errors indicates that teachers and tutors need to emphasize reading comprehension and caution for the math section of the SAT. Instructing students to engage in protocol analysis and then playing their verbalizations back to them and dissecting their thought processes should serve to improve metacognition during the task as well as overall performance (Ericsson, 2006).

The present study has multiple facets and implications and should be replicated with a larger, more controlled design and implementation. The design of a future study of this type should include many more participants with a more diverse socioeconomic background. Past experience with SAT preparation of any kind, including taking SAT exams in the past, should be noted for each participant and controlled for in analysis. The curriculum for participants should be standardized beforehand and regulated throughout the tutoring process such that participants are moving at the same pace and are encountering the same assignments. The baseline and anchor questionnaires had sufficient internal reliabilities (see Table 4) to suggest that they could be used in future research, although presentation of the anchor questionnaire on multiple occasions throughout the tutoring process is suggested. Diagnostic tests involving protocol analysis are an invaluable tool to determine participants' thought processes during assessment, and the inclusion of an ESM measure during the task would give on-line subjective experience data that could be compared to data from the Sloan Study (Shernoff et al., 2003).

The current study provides a framework for the assessment of students' subjective experiences while engaging in SAT tutoring. The use of ESM questions before, during, and after tutoring can give insight into the opinions of students with regards to tutoring and can provide suggestions for producing optimal student engagement and motivation. The current study also implemented protocol analysis using recording technology that is readily available and easy to use and thus can be integrated into tutoring curricula smoothly and with strong performance implications. Systematic assessment of the types of errors that students make while working on SAT problems helps tutors know exactly what aspect of the assignment a student is having trouble with, and can streamline tutoring sessions to encourage maximal allocation of time. Lastly, evaluating the dispositional and practice performance variables that produce the greatest improvement on the SAT can help teachers, parents, and tutors know what is needed to produce the best performance in students.

References

Ajzen, I. (2002). Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of Applied Social Psychology, 32*, 665-683.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84,* 191-215.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.

Burger, J. M. (1989). Negative reactions to increases in perceived personal control. *Journal of Personality and Social Psychology, 56,* 246-256.

Charness, N., Tuffiash, M., Krampe, R., Reingold, E., & Vasyukova, E. (2005). The role of deliberate practice in chess expertise. *Applied Cognitive Psychology, 19,* 151–165.

College Board. (2010). *Retaking the SAT*. http://professionals.collegeboard.com/testing/sat-reasoning/scores/retake.

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. San Francisco: Jossey-Bass.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience.* New York: HarperPerennial.

Csikszentmihalyi, M. (1997). *Finding Flow*. New York: Basic Books.

Csikszentmihalyi, M., & Csikszentmihalyi, L. S. (Eds) (1988). *Optimal Experience: Psychological Studies of Flow in Consciousness*. New York: Cambridge University Press.

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *Journal of Nervous and Mental Disease. 175*, 526-536.

Csikszentmihalyi, M., and Schneider, B. (2000). *Becoming adult: How teenagers prepare for the world of work*. New York, Basic Books.

Duckworth, A. L., Peterson, C., Matthews, M. D., Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92, 1087–1101.

Elbrink, M. (2008). Analyzing and addressing common mathematical errors in secondary education. *B.S. Undergraduate Mathematics Exchange, 5*.

Ericsson, K.A. (2003). The acquisition of expert performance as problem solving. In J.E.Davidson & R.J.Sternberg (Eds.), *The psychology of problem solving* (pp. 31–83). Cambridge, UK: Cambridge University Press.

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine, 79,* S70-81.

Ericsson, K. A. (2006). Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative task. In K. A. Ericsson, N. Charness, P. Feltovich, and R. R. Hoffman, R. R. (Eds.). *Cambridge handbook of expertise and expert performance* (pp. 223-242). Cambridge, UK: Cambridge University Press.

Ericsson, K. A., Krampe, R. T., & Tesche-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*, 363-406.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol Analysis*. Cambridge, MA: MIT Press.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data (revised edition).* Cambridge, MA: Bradford books/MIT Press.

Finn, J. D., Rock, D. A. (1997). Academic success among students at risk for school failure. *Journal of Applied Psychology, 82*, 221–234.

Hektner, J., & K. Asakawa. (2000). Learning to like challenges. *Becoming adult: How teenagers prepare for the world*. M. Csikszentmihalyi and B. Schneider. New York, Basic Books**:** 95-112.

Jackson, L. E., & Coursey, R. D. (1988). The relationship of God control and internal locus of control to intrinsic religious motivation, coping, and purpose in life. *Journal for the Scientific Study of Religion, 27(3),* 399-410.

Jackson, S. A., & Csikszentmihalyi, M. (1999). *Flow in sport.* Champaign, IL: Human Kinetics.

John, O. P., & Srivastava, S. (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin, & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). New York: Guilford.

Joseph, M. W. (2004). *A Detailed and Comprehensive Operationalization of SAT Coaching and Analysis of Efficacy* (Doctoral dissertation). Temple University, Philadelphia, PA.

Judge, T.A., Locke, E.A., & Durham, C.C. (1997). The dispositional causes of job satisfaction: A core evaluations approach. *Research in Organizational Behavior, 19*, 151-188.

Krampe, R. T., & Ericsson, K. A. (1991). Maintaining excellence: deliberate practice and elite performance in young and older pianists. *Journal of Experimental Psychology*, *125*, 331–59.

Messick, S. (1980). *The effectiveness of coaching for the SAT: Review and reanalysis of research from the fifties to the FTC*. Princeton, NJ: Educational Testing Service.

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38,* 30-38.

Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology, 93*, 116-130.

Pajares, F., & Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemporary Educational Psychology*, 20, 426-443.

Powers, D. E., & Rock, D. A. (1998). Effect of coaching on SAT I: Reasoning scores. College Board Report No. 98-6, New York.

Reynolds, A., Oberman, G. & Perlman, C. (1988). An analysis of a PSAT coaching program for urban gifted students. *Journal of Educational Research Vol. 81 No. 3,* 155-164.

Ross, L. *(*1977*).* The intuitive psychologist and his shortcomings: Distortions in the attribution process*.* In L. Berkowtiz (Ed.), *Advances in experimental social psychology, 10,* 174-221.

Shernoff, D.J.*,* Csikszentmihalyi, M.*,* Schneider, B.*, &* Shernoff, E.S. *(* 2003*).* Student engagement in high school classrooms from the perspective of flow theory. *School Psychology Quarterly,* 18*,* 158–176.

Singh, K., Granville, M., Dika, S. (2002). Mathematics and science achievement: Effects of motivation, interest, and academic engagement. *Journal of Educational Research 95,* 323–332.

Smyth, F. L. (1989). Commercial coaching and SAT scores: The effects on college preparatory students in private schools. *The Journal of College Admissions, 123,* 2-7.

Sternberg, R. J. (1998). Metacognition, abilities, and developing expertise: What makes an expert student? *Instructional Science, 26*, 127-140.

Tangney J.P., Baumeister R.F., Boone A.L. *(2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. Journal of Personality, 72, 271–322.*

Tavani, C. M., & Losh, S. C. (2003). Motivation, self-confidence, and expectations as predictors of the academic performance of our high school students. *Child Study Journal 33*, 141-151.

Ward, W. C., & Jenkins, H. M. (1965). The display of information and the judgment of contingency. *Canadian Journal of Psychology, 19*, 231-241.

Appendix A

Child Baseline Questionnaire

Appendix B

Parent Baseline Questionnaire

Appendix C

Anchor Questionnaire

Appendix D

Think Aloud Instructions

Appendix E

Ratio Diagnostic Test A

Appendix F

Ratio Diagnostic Test B

Appendix G

Average Diagnostic Test A

Appendix H

Average Diagnostic Test B

Table 2

Cohen's d and Within-Samples T Test Results for Anchor Questionnaire Activities

Variable	Activity Type	*M*	*SD*	*d*	*p*
Enjoyment	SAT	3.57	1.36	-2.05	< .001
	Other	5.71	.58		
General Goals	SAT	7.65	1.41	2.41	< .001
	Other	4.89	.80		
Concentration	SAT	7.53	1.46	1.47	< .001
	Other	5.73	1.06		
Challenge	SAT	5.63	1.45	2.07	< .001
	Other	3.17	.85		
Interest	SAT	4.22	1.81	-1.27	< .001
	Other	5.95	.65		
Control	SAT	5.78	1.89	-0.43	.14
	Other	6.45	1.14		
SAT Improvement	SAT	8.18	1.97	3.06	< .001
	Other	3.15	1.24		
Skill	SAT	6.02	1.75	0.10	.72
	Other	5.86	1.32		

Table 3

Descriptive Statistics and Correlations for Diagnostic Test Measures

Variable	1	2	3	4	5	6	7
1. Misread errors (%)	-	.47*	.12	-.08	.30	.50**	.11
2. Computation errors (%)		-	.28	.24	.19	.58**	.05
3. Conceptual errors (%)			-	.69***	.42*	.86***	.10
4. Skipped questions (%)				-	-.22	.60**	.57**
5. Miscellaneous errors (%)					-	.54**	-.62***
6. Total errors (%)						-	.09
7. Improvement pre-post (%)							-
Mean (%)	7.75	3.03	6.77	3.03	4.02	24.90	6.08
SD (%)	4.33	2.30	4.66	5.72	4.17	13.42	7.84

* $p < .10$. ** $p < .05$. *** $p < .01$.

Note. Means do not add to the total because unintelligible responses were removed from analysis.

Table 4

Descriptive Statistics and Correlations for Gender, Academic, and Anchor Questionnaire Measures

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1. Female	-	.09	.49**	.21	.03	.33	.28	.14	-.37	.27	.30	-.29
2. GPA[a]		-	.55**	.52**	.28	.24	.24	-.17	.08	.25	.02	.44*
3. First Practice Test Math Score			-	.74***	.36	.32	.19	-.14	.14	.65***	.16	.17
4. Final Practice Test Math Score				-	.28	.05	.06	-.45*	.18	.49**	.25	.29
SAT Preparation Experience												
5. Enjoyment					-	.34	.43*	-.33	.72***	.72***	.53**	.63***
6. Importance to Goals						-	.74***	.38	.09	.35	.41	-.003
7. Concentration							-	.17	.23	.31	.16	.38
8. Challenge								-	-.22	-.40	-.10	-.63***
9. Interest									-	.45*	.32	.54**
10. Control										-	.46*	.45*
11. Relevance to SAT improvement											-	-.02
12. Skill												-
Cronbach's Alpha					.83	.82	.86	.71	.92	.88	.99	.90
Mean	41%	3.6	667.06	685.29	3.57	7.65	7.53	5.63	4.22	5.78	8.18	6.02
SD		0.35	77.6	73.24	1.36	1.41	1.46	1.45	1.81	1.89	1.97	1.75

* $p < .10$. ** $p < .05$. *** $p < .01$.

[a] n=16 (GPA was not reported by one participant).

Table 5

Correlations Between Gender, Academic and Anchor Questionnaire Measures and Diagnostic Test Errors

Variable	Misread errors (%)	Computation errors (%)	Conceptual errors (%)	Skipped questions (%)	Miscellaneous errors (%)	Total errors (%)	Improvement pre-post (%)
Female	-.55**	-.16	-.06	.01	-.34	-.35	.17
GPA[a]	-.26	-.31	-.53**	-.23	-.46*	-.58**	.27
First Practice Test Math Score	-.29	-.03	-.74***	-.37	-.65***	-.74***	.21
Final Practice Test Math Score	-.39	-.16	-.67***	-.19	-.71***	-.71***	.12
Enjoyment	.01	-.11	-.60**	-.35	-.09	-.38	.07
Importance to Goals	-.14	-.02	-.37	-.32	-.09	-.35	.13
Concentration	-.28	.02	-.29	.05	-.18	-.23	.27
Challenge	.14	.32	.27	.02	.10	.22	.02
Interest	.01	.12	-.31	-.02	-.07	-.10	.09
Control	-.27	-.22	-.69***	-.30	-.36	-.60**	.09
Relevance to SAT improvement	-.35	-.48*	-.29	-.26	-.10	-.43*	-.12
Skill	.03	-.07	-.43*	-.01	-.09	-.17	.11

* $p < .10$. ** $p < .05$. *** $p < .01$.

Table 6

Summary of Hierarchical Regression Analysis for Variables Predicting Final Practice Test Math Score

Variable	Model 1	Model 2	Model 3
	β	β	β
First Practice Score	.74***	.69***	.70***
Challenge		-.36**	-.35*
Diagnostic Improvement			-.02
Change in R^2	.55***	.12**	.001

* $p < .10$. ** $p < .05$. *** $p < .01$.