Some Shortcomings of Soliciting Students' Self-Reported SAT Scores

Richard E. Mayer, Andrew T. Stull, Julie Campbell, Kevin Almeroth, Bruce Bimber, Dorothy Chun, and Alan Knight
University of California, Santa Barbara

Abstract

The authors analyzed self-reported SAT scores and actual SAT scores for 285 college students. Students overestimated their actual SAT scores by an average of 23 points (*SD* = 65, *d* = .72), with 7% under-reporting, 53% reporting accurately, and 40% over-reporting, indicating a systematic bias towards over-reporting. The amount of over-reporting was greater for lower-scoring than higher-scoring students, was greater for upper division than lower division students, and was equivalent for men and women. There was a strong correlation between self-reported and actual SAT scores (*r* = .87), indicating high validity of students' memories of their scores. Results are consistent with a motivated distortion hypothesis. Caution is suggested in using self-reported SAT scores in educational research.

Objectives and Theoretical Framework

Self-reported SAT scores are commonly used in educational research involving college students, sometimes to describe the characteristics of the sample, sometimes to use achievement as a main factor in a study, and sometimes to statistically control for the effects of achievement. In spite of the fact that self-reported SAT scores are sometimes solicited in educational research, little is known concerning the accuracy of self-reported SAT scores (Kuncel, Credé, & Thomas, 2005). In a recent review, Kuncel, Credé, and Thomas (2005) summarized the results of four studies involving a total of 292 participants in which 12% under-reported, 36% accurately reported, and 55% over-reported their SAT scores. These results suggest an over-reporting bias, although the size of the bias in terms of points was not provided. Given the relatively low number of participants, and that some of the research reports are not generally available, the current study provides useful new information based on a sample of 285 students--effectively doubling the total number of participants reported in the literature.

We explore three possible cognitive mechanisms underlying the accuracy of self-reported SAT scores: accurate reporting, memory deterioration, and motivated distortion. First, according to the motivated distortion view, people have valid memories of their SAT scores but distort them in systematic ways to protect their self-esteem. This scenario would be reflected in good performance on measures of validity (such as a high correlation between reported and actual SAT score) and poor performance on measures of bias (such as a high mean difference between reported minus actual scores and a high difference between the number of over- and under-reporters). Second, according to the memory deterioration view, memories for SAT score could be inaccurate because of general deterioration of the memory trace, and then reported without bias. This scenario would be reflected in poor performance on measures of validity and low bias. Third, according to the accurate reporting view, memories for SAT score should be accurate, especially given their importance to entering college students, and reported without bias, given the non-consequential nature of the task. This scenario would be reflected in good performance on measures of validity and low bias. We tested these predictions for all students as well as for several subgroups: low achieving versus high achieving students, lower division versus upper division students, and men versus women.

Data Source

The participants were 285 students at the University of California, Santa Barbara who reported their SAT scores on a questionnaire and for whom actual SAT scores were available. The mean actual SAT score was 1219 (SD = 130). The participants consisted of 39 students in an upper division computer science course, 173 students in a lower division history course, and 73 students in an upper division psychology course. There were 133 women and 152 men; there were 158 freshmen and sophomores (i.e., lower division students) and 127 juniors and seniors (i.e., upper division students). For purposes of subsequent analyses, we also partitioned the students into a higher-achieving subgroup (i.e., with actual SAT scores at or above the mean for the sample) and a lower-achieving subgroup (i.e., with actual SAT scores below the mean).

Method

Materials. The materials consisted of a consent form and a 3-page questionnaire, each printed on 8.5 x 11 inch (21.25 x 27.50 cm) sheets of paper. The consent form

explained that participation was voluntary and that all responses would be confidential. The questionnaire contained 50 questions mainly about learning preferences and experiences including a request for students to write down SAT scores. The consent form and questionnaire were stapled together in packets, with each packet having a unique code number printed in the upper right corner.

Procedure. A member of the research team distributed a packet containing the consent form and questionnaire in class during the first week of the quarter in an upper division computer science course, a lower division history course, and an upper division psychology course. The researcher described the project, and then explained that participation was voluntary and that all responses would be confidential. Participants were told that the instructor would not have access to their responses and that all data would be recorded by code number rather than by their name. Participants completed the consent form and questionnaire in class at their own pace. The registrar's office provided demographic information, including actual SAT scores, for each participating student. To insure confidentiality, all data were recorded based on a code number rather than by the student's name.

Results

Evidence for validity: Students' reported SAT scores are strongly related to their actual SAT scores. Overall, there was a strong positive relation between reported and actual SAT scores. The correlation between actual and reported SAT score was $r = .87$, ($n = 285$, $p < .001$), indicating that 76% of the variance in reported SAT scores is accounted for by the variance in the actual SAT score. This effect size is considered large (Cohen, 1988). This finding is consistent with the idea that when students report their SAT scores, they must have a strong memory of their actual SAT score--i.e., students report SAT scores that are grounded in the memory of their actual SAT scores.

Evidence for bias: Students tend to over-report their SAT scores. The mean actual SAT score was 1219 ($SD = 130$) and the mean reported SAT score was 1242 ($SD = 125$). Overall, students over-reported their SAT scores by a mean of 23 points ($SD = 65$, n = 285). Based on a *t*-test, this mean difference between reported minus actual SAT score was significantly greater than zero, $t(284) = 6.052$, $p < .001$. The effect size is $d = .72$, which is in the medium-to-large range (Cohen, 1988). The number of participants who under-reported their SAT score was 20 (7%), the number who accurately reported their SAT score was 151 (53%), and the number who over-reported their SAT score was 114 (40%). The proportion of over-reporters compared to the proportion of under-reporters is significantly greater than chance, $z = 8.103$, $n = 134$, $p < .001$. Students show an overestimation bias that is both statistically significant based on a *t*-test and practically significant based on an effect size analysis. These findings are consistent with the idea that students tend to show systematic bias in the way that report their SAT scores, that is they tend to report higher scores than they actually obtained.

In summary, although actual and reported SAT scores are related to one another (as indicated by a high correlation between actual and reported SAT scores), students tend to be systematically biased in their reporting of SAT scores.

Lower achieving students tend to overestimate more than do higher achieving students. The mean overestimation score for lower achieving students ($M = 38$, $SD = 87$, $n = 136$) was significantly greater than the mean overestimation score for higher achieving

students (*M* = 10, *SD* = 28, *n* = 149), *t*(161) =3.58, *p* < .001. The effect size is *d* = .56 which is considered a medium effect. Overall, there is evidence that lower achieving students overestimate their SAT scores more than do higher achieving students, suggesting that scores get better for those most needing a score boost--at least when students are asked to self-report.

Upper-division students tend to overestimate more than do lower-division students. The mean overestimation score for juniors and seniors (*M* = 34, *SD* = 81, *n* = 127) was significantly greater than the mean overestimation score for freshmen and sophomores (*M* = 15, *SD* = 47, *n* = 158), *t*(193) = 2.316, *p* = .022. The effect size is d = .33 which is considered a small-to-medium effect. Overall, there is evidence that more experienced students overestimate their SAT scores more than do less experienced students. As students spend more time in college--and presumably more time elapses since taking the SAT--their tendency to overestimate their SAT score increases, suggesting that memory fades in a favorable direction. Interestingly, scores get better with time--or at least students' reported scores get better with time.

Men and women do not differ significantly in their tendencies to over-report. The mean over-reporting score for men (*M* = 20, *SD* = 62, *n* = 152) did not differ significantly from the mean over-reporting score for women (*M* = 27, *SD* = 69, *n* = 133), *t*(283) = .86, *p* = *ns*. The effect size is *d* = .10 which is negligible. Overall, there is no evidence that men and women differed in their tendency to over-report. This finding is particularly interesting, given that studies have shown that SAT scores tend to underestimate performance for women and overestimate for men (Halpern 2000; Zwick, 2002).

Educational and Scientific Significance

Theoretical implications. The pattern of results is consistent with the motivated distortion hypothesis, which predicts a high correlation between actual and reported SAT scores (which was obtained) and high amounts of over-reporting (which was obtained); inconsistent with the memory deterioration hypothesis, which predicts a low correlation between actual and reported SAT scores (which was not obtained) and low amounts of over- or under-reporting (which was not obtained); and inconsistent with the accurate memory hypothesis, which predicts a high correlation between actual and reported SAT scores (which was obtained) and low amounts of over- or under-reporting (which was not obtained). Overall, the pattern of results is most consistent with the idea that reported SAT scores are largely based on actual SAT scores--suggesting some validity of self-reported scores--but reported SAT scores tend to be greater than actual SAT scores--suggesting some systematic bias in how students mentally distort or report their scores.

The finding that over-reporting bias is greater for lower-achieving than for higher-achieving students is also consistent with the motivated distortion hypothesis, because lower-achieving students have a greater need to protect their self-image. The finding that over-reporting bias is greater for upper-division than for lower-division students is somewhat consistent with the motivated distortion hypothesis, because upper-division students may be less motivated to accurately report their scores. Finally, we did not expect any sex differences in bias and we did not observe any.

Practical implications. Based on the evidence for systemic bias in self-reported SAT scores (in which students over-reported by 23 points and in which the number of over-reporters was almost 6 times greater than the number of under-reporters), we recommend

caution in using self-reported SAT scores in situations where the student's absolute score is the focus. When there is a need for accurate scores, such as comparing the research sample to standardized norms, self-reported SAT scores are likely to lead to serious problems because of their tendency to be systemically biased.

In contrast, based on the evidence for strong validity of self-reported SAT scores (in which actual SAT scores accounted for 76% of the variance in reported SAT scores), we note that self-reported scores may be useful in situations in which relative differences among students is the main focus. For example, to the extent that actual and reported scores are highly correlated, self-reported SAT scores can be appropriate when the objective is to use SAT scores to partition students into groups or as a covariate.

Finally, in a previous review, Kuncel, Credé, and Thomas (2005, p. 67) noted, "what remains unclear is whether gender or other demographic or individual differences variables are related to different amounts of over-reporting or under-reporting". The current study provides some evidence that researchers should be particularly cautious in using self-reported SAT scores when their sample consists mainly of lower-achieving or more experienced students.

References

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences, Second edition.* Mahwah, NJ: Erlbaum.

Halpern, D. F. (2000). *Sex differences in cognitive abilities (3rd ed*). Mahwah, NJ: Erlbaum.

Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research, 75*, 63-82.

Zwick, R. (2002). *Fair game? The use of standardized admissions tests in higher education.* New York: Routledge Falmer.