Educ Psychol Rev (2007) 19:443–454
DOI 10.1007/s10648-006-9034-z

ORIGINAL ARTICLE

Overestimation Bias in Self-reported SAT Scores

Richard E. Mayer · Andrew T. Stull · Julie Campbell · Kevin Almeroth · Bruce Bimber · Dorothy Chun · Allan Knight

Published online: 8 December 2006

Abstract The authors analyzed self-reported SAT scores and actual SAT scores for five different samples of college students (N=650). Students overestimated their actual SAT scores by an average of 25 points (SD=81, d=0.31), with 10% under-reporting, 51% reporting accurately, and 39% 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=0.82), indicating high validity of students' memories of their scores. Results replicate previous findings (Kuncel, Credé, & Thomas, 2005) and are consistent with a motivated distortion hypothesis. Caution is suggested in using self-reported SAT scores in psychological research.

Keywords Self-report · SAT · Research methodology

Introduction

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 (Kuncel *et al.*, 2005). In this study, we are concerned with a practical question, "How accurate are self-reported SAT scores?" and a theoretical question, "What are the cognitive mechanisms underlying the accuracy of self-reported SAT scores?"

This research was supported by a grant from the Andrew W. Mellon Foundation.

R. E. Mayer · A. T. Stull · J. Campbell · K. Almeroth · B. Bimber · D. Chun · A. Knight
University of California, Santa Barbara, USA

R. E. Mayer (✉)
Department of Psychology, University of California, Santa Barbara, CA 93106, USA
e-mail: mayer@psych.ucsb.edu

Practical question: How accurate are self-reported SAT scores?

In a recent review, Kuncel *et al.* (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 a replication addressing this practical question based on five samples totaling 650 students—effectively tripling the total number of participants reported in the literature. In addition, in the current study we examine two indices of bias—the percentage of students who under-report, accurately report, and over-report (which was reported by Kuncel *et al.*) and the mean difference between reported minus actual SAT score (which was not reported by Kuncel *et al.*). In the current study, students were asked to report their SAT scores in a format that ensured their complete anonymity, so there is no obvious reason to intentionally distort their reporting.

Kuncel *et al.* (2005) also noted the need for future research to examine bias in various subgroups. In compliance, the current study examines the extent of bias in various subgroups including lower- and higher-achieving students, lower division and upper division students, and men and women. On a practical level, it is useful to determine the extent of bias in college students' self-reported SAT scores.

In addition, Kuncel *et al.* (2005) summarized the results of six studies involving a total of 719 participants in which the sample size weighed mean observed correlation between actual and reported SAT score was $r=0.82$. This result suggests strong validity of self-reported SAT scores, in that 67% of the variance in reported SAT scores can be accounted for by the variance in actual SAT scores. However, the specific studies were not identified so it is not possible to review the method used to generate the results. The present study adds to the existing literature by examining the correlation between actual and reported SAT score in a study with a large sample size, which can be broken into subgroups (i.e., lower- and higher-achieving students, lower division and upper division students, and men and women). On a practical level, it is useful to determine the validity of college students' self-reported SAT scores.

Theoretical question: What are the cognitive mechanisms underlying the accuracy of self-reported SAT scores?

As shown in Table I, we examined two dimensions of the data—bias (i.e., the degree to which students systematically overestimate or underestimate their SAT scores) and validity (i.e., the degree to which reported SAT scores are related to actual SAT scores). As one way to measure bias we computed the mean difference between reported SAT score minus actual SAT score (i.e., difference): a large difference indicates a high bias, whereas a negligible difference indicates a low bias. As another way to measure bias, we tallied the number of students whose reported score was below their actual score and the number of students whose reported score was above their actual (i.e., proportion): a large difference between the proportion of over-estimaters and under-estimaters indicates high bias, and a low difference between the proportion of over-estimaters and under-estimaters indicates low bias. To measure validity we computed the correlation between reported and actual SAT score (i.e., correlation): a high correlation indicates high validity, whereas a low correlation indicates low validity.

Table I Bias and Validity Scores

Type of score	Description	High	Low
Bias score			
Difference	Mean difference between reported score minus actual score	Mean difference is large	Mean difference is small
Proportion	Proportion of over-estimaters and under-estimaters	Large difference in proportion of over- and under-estimaters	Small difference in proportion of over- and under-estimaters
Validity score			
Correlation	Correlation between reported and actual scores	Strong correlation	Weak correlation

As shown in first column of Table II, the data we collected allow us to explore three possible patterns of results: high bias/high validity, low bias/low validity, low bias/high validity. First, consider the practical implications of each of these three possible patterns of results, as shown in the second column of Table II. When students display high bias and high validity: (a) caution is advised in using their reported scores to describe the participants because the absolute score is likely to be overestimated, (b) the reported scores can be used as a covariate or to partition the students into groups based on relative differences because each student's reported scores retains to a large extent the same relative position in the distribution as for the actual scores. When the students display low bias and low validity: (a) caution is advised in using the reported scores to describe the participants because the low validity indicates students reported scores are not closely related to their actual scores, and (b) caution is advised in using the reported scores as a covariate or to partition the groups because each student's reported score does not retain the same relative position in the distribution as for actual scores. When students display low bias and high validity, the reported scores can be used for all purposes because they are accurate.

Second consider the theoretical implications of each of the three patterns of results, as summarized in the third column of Table II. A pattern of high bias and high validity suggests that (a) students have some memory for their actual scores because a high percentage of the variance in their reported score is determined by their actual score, and (b) students systematically distort their scores in one direction such as indicated by large tendencies to overestimate. We refer to this interpretation as the *memory distortion view*—i.e.,

Table II Practical and Theoretical Implications of Three Patterns of Results

Pattern of results	Practical implications	Theoretical implications
High bias and high validity	Caution with using reported scores to describe people OK to use reported scores as covariates or to partition	Motivated distortion view
Low bias and low validity	Caution with using reported scores to describe people Caution with using reported scores as covariates or to partition	Memory deterioration view partition
Low bias and high validity	OK to use reported scores to describe people OK to use reported scores as covariates or to partition	Accurate reporting view

people have valid memories of their SAT scores but distort them in systematic ways, perhaps to protect their self-esteem.

A pattern of low bias and low validity suggests that (a) students do not have a strong memory for their actual scores because a low percentage of variance in reported scores is attributable to their actual scores, and (b) students cannot systematically distort because they do not remember their actual scores. We refer to this interpretation as the *memory deterioration view*—memories deteriorate over time without systematic bias in reporting.

Finally, low bias and high validity are the hallmarks of highly accurate performance. We refer to this interpretation of student performance as the *accurate reporting view*—which assumes that people have nearly perfect memory for important facts and that they report their memories without bias. SAT scores are a monumental factor in college students' lives so it is plausible students would have accurate memories of their scores.

We examined the pattern of bias and validity scores for five samples of students as well as for several subgroups for which we have data: low achieving versus high achieving students, lower division versus upper division students, and men versus women. We defined low achieving students as those scoring below the mean of the sample on the SAT, and high achieving students as those scoring at or above the mean on the SAT. We chose to examine high and low achieving students separately in order to determine whether students with lower SAT scores would be more likely to compensate by over-reporting their scores. We defined lower division students as those officially designated as freshmen or sophomores based on their completed credit units, and we defined upper division students as those designated as juniors or seniors. Finally, concerning men and women, we have no reason to suspect differences in these predicted patterns attributable to the student's gender, but we include gender as a factor because sex differences in SAT scores have received much attention in the literature (Halpern, 2000; Zwick, 2002). In particular, Zwick (2002) reports that SAT scores underpredict the academic performance of women and overpredict the academic performance of men (e.g., on average, women obtain higher grades than men with the same SAT score), so there might be some expectation that women show more bias in self-reported scores than men.

Materials and Methods

Participants

The participants were five samples of college students at a selective university in California who reported their SAT scores on a confidential questionnaire and for whom actual SAT scores were available. The participants consisted of 78 students in a freshman writing course (i.e, Writing), 200 students in a lower division history course in American History (i.e., American History), 168 students in a lower division history course in Western Civilization (i.e., Western Civilization), 141 students in an upper division psychology course in Educational Psychology (i.e. Educational Psychology), and 63 students in an upper division computer science course in Networking (i.e., Computer Science). Table III lists the mean actual score (and standard deviation) on the combined SAT, SAT-Verbal, and SAT-Math, the percentage of men and women, and the percentage of freshmen, sophomores, juniors and seniors, for each of the five samples of college students. The samples in Table III are arranged in order of percentage of freshmen and sophomores, with Writing, American History, and Western Civilization containing mainly lower division

Table III Characteristics of Five Samples of College Students

Sample	Number	Combined SAT		SAT-Verbal		SAT-Math		Gender (%)		Year (%)			
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*F*	*Fr*	*So*	*Jr*	*Sr*
Writing	78	1150	122	552	69	598	80	49	51	90	10	0	0
American History	200	1226	137	609	84	617	79	58	42	57	33	6	4
Western Civilization	168	1189	142	585	85	604	75	40	60	49	29	17	5
Computer Science	63	1282	131	596	101	686	58	87	13	0	0	19	81
Educational Psychology	141	1170	116	572	69	598	73	21	79	0	1	28	70
Total	650	1201	137	587	83	614	79	47	53	41	19	14	26

students (i.e., freshmen and sophomores) and with Educational Psychology and Computer Science containing mainly upper division students (i.e., juniors and seniors).

Based on data for all 650 students in the five samples, the mean actual SAT score was 1201 (SD=137), there were 307 men and 343 women, and there were 266 freshmen, 124 sophomores, 92 juniors, and 168 seniors. For purposes of subsequent analyses, we also partitioned the students into a higher-achieving subgroup (i.e., 327 students with actual SAT scores at or above the mean for the sample) and a lower-achieving subgroup (i.e., 323 students with actual SAT scores below the mean).

Materials and methods

The materials consisted of a consent form and a three-page questionnaire, each printed on 8.5×11 in. (21.25×27.50 cm) sheets of paper. The consent form explained that participation was voluntary, that all responses would be confidential, and that the course instructor would not have access to any of the data. The consent form stated "we will record the data by code number rather than by your name." The questionnaire solicited demographic information and information about the student's preferred learning style, but did not ask for the student's name. The questionnaire contained 50 questions 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 as part of a larger project. The researcher described the project, and then explained that participation was voluntary and that all responses would be confidential. The researcher showed that the top sheet (containing the consent form) was the only form with the participant's name, and that it would be separated from the questionnaire so all student responses would be stored only by code number. 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. When a participant handed in the packet containing the consent form and questionnaire, the researcher separated the consent form (containing the participant's name) from the questionnaire (which contained a code number but no name). Data were collected during the 2004–2005 and 2005–2006 academic years. The registrar's office provided demographic information, including actual SAT scores,

for each participating student. We removed the participants' names and substituted their code number (which was taken from the top sheet of their consent form/questionnaire packet). To ensure confidentiality, all data were recorded based on a code number rather than the student's name, and the sheets linking the students' names and code numbers were destroyed.

Results

Are self-reported SAT scores biased and valid?

The first step is to determine whether students are biased in reporting their SAT scores (i.e., whether they systematically overestimate or underestimate their scores when they are asked to self-report) and whether students are valid in reporting their SAT scores (i.e., whether their self-reported scores are systemically related to their actual scores).

Evidence for bias: Students tend to over-report their SAT scores

The left side of Table IV reports the mean overestimation of combined SAT score for students in each of the five samples and for all students combined. For each sample, the self-reported score was greater than the actual score on average, with mean overestimation ranging from 14.6 to 35.3 points.[1] We computed a difference score for each student by subtracting the actual SAT score from the reported SAT score. For each sample, *t*-tests revealed that the difference between reported and actual score was significantly different than zero, and effect sizes ranged from d=0.25 to d=0.45. Based on the combined data, the mean overestimation was 25.2 points. The bottom line of Table IV shows that this difference was statistically significant based on a *t*-test, and that the effect size of d=0.31 was in the small to medium range.

The left side of Table V reports the mean overestimation of SAT-Verbal score for students in each of the five samples and for all students combined. For each sample, the self-reported score was greater than the actual score on average, with mean overestimation ranging from 3.7 to 21.8 points. For three of the five samples, *t*-tests revealed that the difference between reported and actual score was significantly different than zero and for two of the five samples no significant difference was found. The effect sizes ranged from d=0.12 to d=0.48. Based on the combined data, the mean overestimation for SAT-verbal score was 16.0 points. The bottom line of Table V shows that this difference was statistically significant based on a *t*-test, and that the effect size of d=0.31 was in the small to medium range.

The left side of Table VI reports the mean overestimation of SAT-Math score for students in each of the five samples and for all students combined. For each sample, the self-reported score was greater than the actual score on average, with mean overestimation ranging from 5.6 to 15.9 points. For four of the five samples, *t*-tests revealed that the difference between reported and actual score was significantly different than zero and for one of the five samples no significant difference was found. The effect sizes ranged from

[1] We note that the educational psychology class showed the greatest overestimation bias. All students in the class are psychology majors, and the psychology major is one of the most competitive on campus. Our goal in this project was not to make systematic comparisons across classes, but further research on this topic is warranted. We also recognize many of the upper division students come from this class, so further research is warranted concerning differences between upper and lower division students.

Table IV Mean Difference between Reported and Actual Score (and Standard Deviation), Proportion of Over- and Under-estimaters, and Correlation between Reported and Actual Scores for Five Samples of College Students on Combined SAT

Sample	Number	Difference score						Proportion				Correlation	
		M	*SD*	*d*	*t*	*df*	*p*	Under	Right	Over	*p*	*r*	*p*
Writing	78	14.6	44.4	0.33	2.91	77	0.005	0.10	0.57	0.33	0.003	0.93	0.000
American History	200	21.7	71.7	0.30	4.27	199	0.000	0.08	0.53	0.39	0.000	0.86	0.000
Western Civilization	168	26.6	108.7	0.25	3.17	167	0.002	0.13	0.51	0.36	0.000	0.72	0.000
Educational Psychology	141	35.3	77.8	0.45	5.39	140	0.000	0.08	0.43	0.49	0.000	0.77	0.000
Computer Science	63	23.0	57.6	0.40	3.17	62	0.002	0.08	0.54	0.38	0.000	0.90	0.000
Total	650	25.2	80.7	0.31	7.96	649	0.000	0.10	0.51	0.39	0.000	0.82	0.000

d=0.10 to d=0.37. Based on the combined data, the mean overestimation for SAT-Math score was 9.2 points. The bottom line of Table VI shows that this difference was statistically significant based on a *t*-test, and that the effect size of d=0.20 was in the small range.

In summary, students showed 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. This pattern replicates and extends findings summarized by Kuncel *et al*. (2005).

More evidence for bias: The proportion of over-estimaters is greater than the proportion of under-estimaters

The middle portion of Table IV shows the proportion of students who under-reported, accurately reported, and over-reported their combined SAT score for each of five samples

Table V Mean Difference between Reported and Actual Score (and Standard Deviation), Proportion of Over- and Under-estimaters, and Correlation between Reported and Actual Scores for Five Samples of College Students on SAT-Verbal

Sample	Number	Difference score						Proportion				Correlation	
		M	*SD*	*d*	*t*	*df*	*p*	Under	Right	Over	*p*	*r*	*p*
Writing	78	3.7	31.3	0.12	1.05	77	ns	0.26	0.45	0.29	ns	0.90	0.000
American History	200	15.3	51.2	0.30	4.22	199	0.000	0.18	0.35	0.37	0.001	0.81	0.000
Western Civilization	168	21.0	63.4	0.33	4.29	167	0.000	0.15	0.43	0.42	0.000	0.72	0.000
Educational Psychology	141	21.8	45.3	0.48	5.71	140	0.000	0.18	0.28	0.54	0.000	0.79	0.000
Computer Science	63	7.1	41.0	0.17	1.38	62	ns	0.22	0.40	0.38	ns	0.92	0.000
Total	650	16.0	51.0	0.31	8.00	649	0.000	0.19	0.40	0.41	0.000	0.81	0.000

Table VI Mean Difference between Reported and Actual Score (and Standard Deviation), Proportion of Over- and Under-estimaters, and Correlation between Reported and Actual Scores for Five Samples of College Students on SAT-Math

Sample	Number	Difference score					Proportion				Correlation		
		M	*SD*	*d*	*t*	*Df*	*p*	Under	Right	Over	*p*	*r*	*p*
Writing	78	10.9	33.6	0.32	2.86	77	0.005	0.17	0.41	0.42	0.005	0.91	0.000
American History	200	6.4	39.9	0.16	2.26	199	0.025	0.21	0.47	0.32	0.064	0.87	0.000
Western Civilization	168	5.6	56.5	0.10	1.29	167	ns	0.26	0.46	0.28	ns	0.74	0.000
Educational Psychology	141	13.5	49.5	0.27	3.24	140	0.001	0.26	0.32	0.43	0.018	0.76	0.000
Computer Science	63	15.9	42.5	0.37	2.96	62	0.004	0.16	0.44	0.40	0.017	0.76	0.000
Total	650	9.2	46.5	0.20	5.04	649	0.000	0.22	0.42	0.36	0.000	0.83	0.000

and for all students combined. For each of the five samples the proportion of over-estimaters was significantly greater than the proportion of under-estimaters, based on binomial tests. When we combined the data across all five samples, the number of over-estimaters was approximately four times greater than the number of under-estimaters, as shown in the bottom line of Table IV. This difference was statistically significant, based on a binomial test.

Similar findings were found for students' reporting of SAT-Verbal and SAT-Math scores, individually. The middle portion of Table V shows the proportion of students who under-reported, accurately reported, and over-reported their SAT-Verbal score for each of five samples and for all students combined. For three of the five samples the proportion of over-estimaters was significantly greater than the proportion of under-estimaters, based on binomial tests. When we combined the data across samples, the number of over-estimaters was approximately two times greater than the number of under-estimaters, as shown in the bottom line of Table V. This difference was statistically significant, based on a binomial test.

The middle portion of Table VI shows the proportion of students who under-reported, accurately reported, and over-reported their SAT-Math score for each of five samples and for all students combined. For three of the five samples the proportion of over-estimaters was significantly greater than the proportion of under-estimaters, based on binomial tests. When we combined the data across samples, the number of over-estimaters was approximately 1.6 times greater than the number of under-estimaters, as shown in the bottom line of Table VI. This difference was statistically significant, based on a binomial test.

These patterns are similar to those reported by Kuncel *et al.* (2005), and are consistent with the mean bias scores reported in the previous subsection. Overall, students demonstrated a systematic bias towards over-reporting their SAT scores.

Evidence for validity: Students' reported SAT scores are strongly related to their actual SAT scores

The right side of Table IV reports the correlation between reported and actual SAT score for each of the five samples and for all students combined. For each sample, there was a strong positive relation between reported and actual SAT score, ranging from r=0.72 to r=0.93. Overall, the correlation between actual and reported SAT score was r=0.82, (n=650, p<

0.001), indicating that 67% of the variance in reported SAT scores is accounted for by the variance in the actual SAT score. All correlations were statistically significant.

The right side of Table V reports the correlation between reported and actual SAT-Verbal score for each of the five samples and for all students combined. For each sample, there was a strong positive relation between reported and actual SAT-Verbal score, ranging from r=0.72 to r=0.92. Overall, the correlation between actual and reported SAT score was r=0.81, (n=650, p<0.001), indicating that 66% of the variance in reported SAT scores is accounted for by the variance in the actual SAT score. All correlations were statistically significant.

The right side of Table VI reports the correlation between reported and actual SAT-Math score for each of the five samples and for all students combined. For each sample, there was a strong positive relation between reported and actual SAT-Math score, ranging from r=0.74 to r=0.91. Overall, the correlation between actual and reported SAT score was r=0.83, (n=650, p<0.001), indicating that 69% of the variance in reported SAT scores was accounted for by the variance in the actual SAT score. All correlations were statistically significant.

The correlations reported in Tables IV, V, and VI are consistent with the idea that when students report their SAT scores, they have a strong memory of their actual SAT score (i.e., students reported SAT scores that are grounded in the memory of their actual SAT scores). This pattern replicates findings summarized by Kuncel *et al.* (2005).

In summary, although actual and reported SAT scores were related to one another (as indicated by a high correlation between actual and reported SAT scores), students tended to be systematically biased in their reporting of SAT scores (as indicated in the previous section). In short, in each of the five samples students displayed a pattern of high bias and high validity.

Who tends to overestimate?

The next step is to determine whether certain types of people overestimate more than other types of people. In this section, we compare higher and lower achieving students, lower division and upper division students, and men and women, based on the combined sample of all 650 students.[2]

Lower achieving students tend to overestimate more than do higher achieving students

The mean difference between reported and actual score on the combined SAT was significantly greater for lower achieving students (M=42.1, SD=104.9, n=327) than for higher achieving students (M=8.0, SD=37.4, n=323), t(648)=5.501, p<0.001. The effect size was d=0.48 which is considered a medium effect.

Similar patterns were found for SAT-Verbal and SAT-Math individually. The mean overestimation of SAT-Verbal score was significantly greater for lower achieving students (M=25.4, SD=63.8, n=327) than for higher achieving students (M=6.5, SD=30.8, n=323), t(648)=4.813, p<0.001. The effect size was d=0.40 which is considered in the small to medium range. The mean overestimation of SAT-Math score was significantly greater for lower achieving students (M=16.7, SD=57.7, n=327) than for higher achieving students

[2] Given that three t-tests were conducted on the same data (partitioned for high versus low SAT score, lower- versus upper-division, and men versus women, respectively), there exists the danger that Type 1 error was inflated. To address this issue, we applied a Bonferroni procedure, which showed that all significant differences reported in the results section remained statistically significant.

(M=1.6, SD=29.3, n=323), t(648)=4.204, p<0.001. The effect size was d=0.35 which is considered in the small to medium range. Overall there is evidence that lower achieving students tended to overestimate their SAT scores to a larger extent than did higher achieving students.

Upper-division students tend to overestimate more than do lower-division students

The mean difference between reported and actual combined SAT was significantly greater for juniors and seniors (M=38.5 SD=86.9, n=260) than for freshmen and sophomores (M=16.3, SD=75.1, n=390), t(648)=3.464, p<0.001. The effect size was d=0.27 which is considered a small-to-medium effect.

Similar patterns were found for SAT-Verbal and SAT-Math individually. The mean difference score on SAT-Verbal was significantly greater for juniors and seniors (M=25.0, SD=58.0, n=260) than for freshmen and sophomores (M=10.0, SD=44.8, n=390), t(648)=3.731, p<0.001. The effect size was d=0.29 which is considered in the small to medium range. The mean difference score on SAT-Math was marginally greater for juniors and seniors (M=13.5, SD=48.8, n=260) than for freshmen and sophomores (M=6.3, SD=44.7, n=390), t(648)=1.916, p=0.056. The effect size was d=0.15 which is considered negligible. Overall, there is evidence that more experienced students overestimated their SAT scores more than did less experienced students.

Men and women do not differ significantly in their tendencies to overestimate

The mean difference score for men (M=22.2, SD=63.5, n=307) did not differ significantly on combined SAT from the mean difference score for women (M=27.9, SD=93.4, n=343), t(648)=0.90, p=*ns*. The effect size was d=0.07 which is negligible. Similar patterns were obtained on SAT-Verbal [t(648)=1.620, p=ns] and SAT-Math [t(648)=0.216, p=ns] individually. Overall, there was 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).

Discussion

Empirical findings

Our primary goal was to conduct a set of replication studies concerning the bias and validity of students' self-reported SAT scores. The main findings of this study were that college students tended to over-report their SAT scores (by a mean of 25 points) but their reported scores correlated strongly with their actual scores (with actual scores accounting for about two-thirds of the variance in reported scores). In short, in each of the five samples, students showed a pattern of high bias and high validity. These findings essentially replicated those reported by Kuncel *et al.* (2005), although our study involved more participants from five different samples, included an additional measure of bias, and added supplemental analyses by subgroups. In particular, lower-achieving students tended to overestimate more than higher-achieving students and upper-division students tended to overestimate more than lower-division students.

Practical implications

First, we offer implications based on the pattern of results for the entire sample. Based on the evidence for systemic bias in self-reported SAT scores (in which students over-reported by 25 points and in which the number of over-reporters was almost four 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 or describing the characteristics of the sample, 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 67% 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 achievement groups or as a covariate.

In short, the practical implication of overestimation bias is that using self-reported SAT scores as the basis for computing a mean SAT score for a sample does not provide an accurate description of the sample. The practical implication of a high correlation between self-reported and actual SAT scores is that SAT scores can be used to partition a sample into relatively higher and lower performing subgroups.

Second, we examine the implications of the secondary analyses based on the subgroups. In a previous review, Kuncel *et al*. (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 over-reporting was greater for lower-achieving than higher-achieving students, and for upper division than lower division students, whereas men and women did not differ. Researchers should be particularly cautious in using self-reported SAT scores when their sample consists mainly of lower-achieving or more experienced students.

Theoretical implications

The pattern of results is consistent with the memory 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). We tentatively accept the motivated distortion hypothesis, but recognize the need for further testing. In particular, the available data could not pinpoint the locus of the distortion—such as, during storage (i.e., the student's memory representation has been changed) or at retrieval (i.e., the distortion occurs when the students reports the score). Similarly, the available data do not help to determine whether the distortion is unconscious or conscious.

The pattern of results is 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). We reject the memory deterioration hypothesis on the basis of the high correlation between actual and reported SAT scores and on the grounds that students displayed a systematic bias in over-reporting their SAT scores. However, we note that it is possible that memory deterioration may play a role for some subsets of students.

The pattern of results is 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). We reject the accurate memory hypothesis on the grounds that students displayed a systematic bias in over-reporting their SAT scores. However, we note that approximately half of the students reported completely accurate combined SAT scores, so future research should focus on the characteristics and cognitive processes of over- and under-reporters. Further research is also needed to determine whether the same pattern of results can be replicated for other tests such as the ACT, GRE, and MCAT.

Overall, the pattern of correlation 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 report their scores.[3] This paper replicates and extends previous evidence for overestimation bias in students' self-reported SAT scores, and provides an example of the value of replication studies in educational psychology.

References

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.

[3] We distinguish between *validity*—the degree to which the reported scores correlate with actual scores—and *bias*—the degree to which reported score is systematically greater than actual score. It is possible for self-reported scores to be valid and biased, as was the case in this study.