

Accuracy of Self-reported SAT and ACT Test Scores: Implications for Research

James S. Cole · Robert M. Gonyea

Received: 23 January 2009 / Published online: 5 November 2009
© Springer Science+Business Media, LLC 2009

Abstract Because it is often impractical or impossible to obtain school transcripts or records on subjects, many researchers rely on college students to accurately self-report their academic record as part of their data collection procedures. The purpose of this study is to investigate the validity and reliability of student self-reported academic performance. As expected the study finds overall validity of self-reported test scores to be high. However, correlations between self-reported and actual SAT scores are significantly lower than correlations for self-reported and actual ACT Composite scores. This study also confirms prior research which found that when students are inaccurate in reporting their scores, a disproportionate number of them over-report their scores. Also consistent with other studies, this study finds that lower achieving students for both tests are much less accurate when reporting their scores.

Keywords Self-report accuracy · Cognitive distortion · Motivated distortion · Higher education research · SAT/ACT test scores

Introduction

Survey research is perhaps the most common data collection method in higher education as institutions ramp up assessment efforts for both improvement and accountability. Given that it is often impractical or impossible to obtain school transcripts or records for each student; many researchers rely on college students to accurately self-report their academic record as part of the research data collection procedures (Kuncel et al. 2005). Many major national surveys also collect self-reported SAT and ACT scores including, but not limited to, the *Beginning College Survey of Student Engagement* (BCSSE) administered by the Center for Postsecondary Research at Indiana University.

J. S. Cole (✉) · R. M. Gonyea
Center for Postsecondary Research, Indiana University, Eigenmann Hall, Suite 419,
Bloomington IN 47406, USA
e-mail: colejs@indiana.edu

Accuracy of self-reported data has been investigated for many years in the higher education literature. The type of self-reported data investigated in this line of research can generally be categorized as two types: attitudinal and factual (Tourangeau et al. 2000). *Attitudinal* data come from self-reports of beliefs or opinions that are generally not verifiable by an outside source (e.g., “How satisfied are you with your academic advising?”). We must rely on the respondent to answer the question truthfully. Attitudinal variables are most often used as predictors of important student outcomes in higher education research, so the predictive validity of the data is the primary concern (e.g., Braxton et al. 1988).

Factual data on the other hand, can be subject to verification (e.g., “What is your GPA?”). While the researcher relies on the respondent to be truthful, self-reported factual data can be verified against external data sources such as records held by the institution. Whereas the primary concern with attitudinal data often is predictive validity, with factual data the primary concern for the researcher is *concurrent* validity where the investigator is interested in the degree that the measure correlates with criteria that assess the same thing (Anastasi and Urbina 1997; Creswell 2008). For instance, Anaya (1999) investigated the accuracy of self-reported GRE scores against institutional records. Overall, Anaya reported that the correlation between self-reported and actual GRE scores were quite high ($r = .94$), but that higher achieving students and females were significantly more accurate when reporting scores. Anaya also found that when students were not accurate, they were more likely to overstate their scores than understate. Similarly, Cabrera et al. (1990) reported that the correlation between self-reported and institution-reported GPA was .75.

Though other higher education studies have also investigated the accuracy of self-reported factual data, there is currently a paucity of research devoted to the investigating the concurrent validity of factual data in higher education research. Thus, the purpose of this study is to investigate the concurrent validity of student self-reported academic performance on the SAT and ACT against institutional records.

Accuracy of Self-Reported Factual Data

All data have measurement error (Anastasi and Urbina 1997). Ideally error will be minimized so that the reported data are as close to their ‘true’ score as possible. However, most researchers acknowledge that self-reported data, with their reliance on accurate and unbiased memories, are exposed to concerns about reliability and validity (Gonyea 2005; Pearson et al. 1994). The quality of self-reported test scores in research is called into question in part because external data are usually not available to verify them. Ideally, measurement errors should be randomly distributed so that the number and magnitude of overestimates equals the number and magnitude of underestimates. Thus, with an increasing number of cases, the mean of this distribution of errors would approach zero and the mean of the reported scores would be equal to the true score. In situations where the errors are not randomly distributed (i.e., the reported scores are consistently either above or below the actual scores) then systematic bias may exist in the measurement (Schmidt et al. 2003). In this case, total measurement error contains two types, random error and systematic bias.

There are various explanations for error in reporting tests scores. However, for this study we focus on two types of distortion that contribute to self-report errors. One is “motivated distortion” (Williard and Gramzow 2008). According to this perspective, errors in self-reported data are systematic and not random. The student is motivated to purposely provide an inaccurate report of the data requested. For example, some researchers speculate that motivated distortion among students may be due to *social*

desirability bias—the desire to revise a response before communicating it to a researcher to protect self-image or inaccurately project an image of academic performance (Beretvas et al. 2002; Nancarrow and Brace 2000; Dobbins et al. 1993; Shepperd 1993). Responders succumbing to social desirability pressure are either expressing a need to present themselves favorably to the researcher (impression management) and/or a need to preserve their self-esteem (ego defense) (Nancarrow and Brace 2000). For instance, a student who receives a relatively low score on the SAT may feel that the score is not an accurate reflection of his or her abilities or are otherwise faulty, and thereby report a higher SAT on the survey. Previous research has shown that actual test scores (not self-reported) correlate higher with earned GPA in college than did the scores the subject believed would have been fairer, but were not accurate (Shepperd 1993). In addition, students with lower actual test scores not only tended to recall their scores with less accuracy, but also to inflated their scores significantly more than students with higher test scores (Shepperd 1993). However, researchers also believe social desirability bias is smaller in paper and online surveys compared with face-to-face interviews, but can be more pronounced on questions that are more intimidating or anxiety-provoking, as with test scores (Tourangeau et al. 2000).

On the other hand, a recent study found that in addition to motivated distortions such as social desirability bias, *reconstructive memory processes* may also explain the inaccuracies. This “cognitive distortion” perspective claims that self-reported error is due to difficulties with encoding experiences in long term memory and/or retrieval of such memories at a later date (Williard and Gramzow 2008; Gramzow and Williard 2006). According to this perspective, errors in self-reported data are non-systematic and, as a result, should be randomly distributed. The cognitive distortion view claims that honest attempts to fill in gaps of information about one’s past can result in exaggerations distinct from being motivated to portray oneself in a more positive way.

Though there is concern about the accuracy of self-reported test scores, the literature reports relatively high correlations between self-reported and actual test scores. For example, Cassady (2001) found the correlations between actual and self-reported SAT scores to be .73 for the verbal test, .89 for the math test, and .88 for the total score. A meta-analysis by Kuncel et al. (2005) of 91 peer-reviewed studies whose samples totaled over 61,000 students reported a relatively high level of validity in student self-reported GPA with an overall correlation of .84. Other studies are consistent with these results (Herman 2003; Frucot and Cook 1994). However, less than perfect correlations indicate at least some attenuation, a lowering of observed correlation due to unreliable data (Creswell 2008). This has led some researchers to conclude that self-reported variables should be used with caution because of the generally unknown amount and source of error in the self-reported data (Kuncel et al. 2005; Mayer et al. 2007).

Regardless of whether the source of the error is due to motivated or cognitive distortion, it is clear across several studies that students with lower scores were more likely to be inaccurate and to overestimate their test scores (Bahrack et al. 1996; Cassady 2001; Kuncel et al. 2005; Shepperd 1993). For instance, the purpose of a recent study by Mayer and his colleagues was to investigate the “cognitive mechanisms underlying the accuracy of self-reported SAT scores” (p. 443). Their study included 650 undergraduate students. These researchers reported that the mean overestimation of the combined SAT score was 25.2 points and was significantly different than zero with a small-medium effect size ($d = .31$). They also reported that the number of over-estimators was four times greater than the number of students who underestimated their SAT scores and that those students who over-reported were significantly more likely to have lower test scores than students

who did not over-report their scores. They conclude that the data supports a motivation distortion hypothesis and essentially replicates the findings by Kuncel et al. (2005).

This study investigates errors in self-reported SAT and ACT scores from both the cognitive and motivated distortion views. From the cognitive distortion perspective, it is expected that a student will self-report their *single* ACT Composite score more accurately than a student who self-reports their *three* SAT scores (Critical Reading, Mathematics, and Writing). That is, it is expected that it there will be more error in recalling three, three-digit subtest scores ranging from 200 to 800 each (SAT scores), than a single number that ranged between 11 and 36 (ACT Composite). Also, from the motivated distortion perspective, it is expected that students with lower actual SAT/ACT scores will be less accurate in reporting their scores, compared to higher achieving students. Finally, this study will investigate whether or not an SAT total score calculated by summing individual self-reported SAT Critical Reading and Mathematics scores is more accurate than each individual self-reported score in these subject areas. This latter analysis is based on the cognitive distortion perspective which claims that errors in reporting are random. By calculating a total score from the self-reported sub-test scores, errors in the each subtest should tend to cancel each other out thereby resulting in reduced error in the total score.

Therefore, three research questions guided this study:

- (1) Overall, is there difference between the accuracy of self-reported ACT scores compared with that of the three SAT tests?
- (2) Is there difference in the accuracy of self-reported ACT and SAT scores based on the achievement level of the student?
- (3) Is there difference in the accuracy of a combined SAT score (the sum of the self-reported Critical Reading and Mathematics scores) compared to the individual self-reported SAT scores?

Method

The self-reported test scores for this study came from the 2007 *Beginning College Survey of Student Engagement* (BCSSE). In 2007, over 67,000 students starting their first year of college at 126 baccalaureate degree-granting institutions completed BCSSE. BCSSE measures entering 5-year students' pre-college academic and co-curricular experiences, as well as their interest in and expectations for participating in educationally purposeful activities during college. BCSSE administration typically takes place during the summer/early fall and is designed to be paired with the administration of the *National Survey of Student Engagement* at the end of the first college year.

On BCSSE, SAT and ACT test scores are collected by asking students to "please write your scores below (as best as you remember)." Students are provided three boxes to provide the SAT Critical Reading, Mathematical Reasoning, and Writing score. Students are also provided one box to provide their ACT composite score (see Fig. 1). For both, the possible (or valid) ranges are given to aid the students' recall efforts. The BCSSE survey does not ask students to report an overall SAT score.

Actual test score data came directly from the institutions where these students were enrolled and that participated in the spring 2008 administration of the *National Survey of Student Engagement* (NSSE). The test scores reported by these institutions are the official scores they have on file for these students. As part of its random sampling administration process, NSSE collects population data on first-year students directly from the participating

Fig. 1 Student self-reported SAT and ACT scores on BCSSE

10 Did you take the SAT and/or ACT?
 Yes No
 If yes, please write your scores below (as best you remember):

SAT (possible range=200-800)			ACT (possible range=1-36)	
Critical Reading	<input type="text"/>	<input type="text"/>	Composite	<input type="text"/>
Mathematical Reasoning	<input type="text"/>	<input type="text"/>		
Writing	<input type="text"/>	<input type="text"/>		

institutions and requests, but does not require precollege achievement test scores to be included in the data. All students who self-reported ACT scores and had scores provided by their institutions were included in the study. In addition, students who self-reported scores for the ‘new’ SAT that includes reading, math, and writing scores and had corresponding test scores reported by their institutions were included in the study. For the SAT, a calculated, combined score was created using the SAT Critical Reading and Mathematics scores, a common practice for researchers and institutions. For example, the College Entrance Examination Board regularly updates their ACT to SAT conversion table using the combined Critical Reading and Mathematics SAT scores (College Board 2008). This combined SAT score has a minimum score of 400 and a maximum score of 1600. Finally, only those students who graduated from high school the previous spring (2007) were included. In total, there were 24,624 first-year students enrolled at 74 institutions that fit the criteria for inclusion in the study. The sample included 53% females, with 64% identified as Caucasian, 9% Black, 5% Hispanic/Latino, 8% Asian, and 13% other or not indicated. Over 85% of these students graduated from a public high school, 12% from a private, religious-affiliated high school, 2% from a private, independent high school, and 1% were either home schooled or reported the ‘other’ option. Self-reported and actual test scores for this sample are given in Table 1.

For this study, three indicators of “accuracy” were used: frequency of error, magnitude of bias, and correlation. The first, *frequency of error*, was the number of occurrences of either under-reporting, over-reporting, or accurately reporting test scores (Williard and Gramzow 2008). However, many students take the SAT and/or ACT more than once (College Board 2008). Therefore, determining which scores were accurate was not as simple as comparing the self-reported with the institution-reported scores. For example,

Table 1 Self-reported and institution-reported test scores

	N	Self-reported		Institution-reported	
		Mean	SD	Mean	SD
SAT critical reading	8605	552.9	92.9	548.1	90.1
SAT mathematics	8825	568.4	95.8	567.5	89.3
SAT writing	8370	549.4	93.5	544.5	88.0
SAT total	8565	1121.2	163.8	1115.7	157.9
ACT composite	16643	24.0	4.3	23.6	4.4

about one-half of SAT test-takers take the exam more than once, thus these students have more than one set of scores to report. The most common test-taking schedule for those who take it multiple times is to sit for the SAT in the spring of the junior year and then once again in the fall of the senior year. Of those who take the exam a second time, about 55% improve their score, while about 35% get lower scores (College Board 2008). Thus, when students were asked to recall their scores, it is not clear which scores they were recalling, nor was it clear in the present study which scores the institution reported to NSSE as the official scores. Multiple test-takers also tend to be higher achieving. In 2008, students who took the exam twice scored on the first exam an average a 25 points higher than students who took the exam only once. According to the College Board, students who took the exam twice improved their individual test scores by an average of 15 points. Thus, for this study, rather than expecting the student to perfectly match the institution-reported score, a somewhat relaxed definition of accuracy was used. An ‘accurate’ self-reported SAT score was defined as within 20 points (plus or minus) of the institution-reported score. Likewise, a proportionally equivalent margin of error based on the possible range of scores for the ACT is one point (plus or minus). This is also a criterion for accuracy used by Anaya (1999) in her study of accuracy of self-reported test scores.

The second indicator of accuracy used for the study was the *magnitude of bias* (Mayer et al. 2007), calculated as the ratio of the difference between the self-reported and institutional scores to the range of possible scores for each test. For the SAT, scores for each of the three sections range 600 points (from 200 to 800), and the ACT Composite score ranges 25 points (from 11 to 36). This magnitude of bias ratio effectively standardized the difference scores so that they could be used for mean comparisons between the SAT and ACT in self-reported test bias. For example, if a student self-reported an SAT Writing score of 590 but his institution score was 510, then his bias ratio would be .13 $((590-510)/600 = .13)$. Likewise, if a student reported an ACT score of 24 but her actual score was 21, then her bias ratio would be .12 $((24-21)/25 = .12)$.

The third indicator of accuracy used in the study was the *correlation* between the self-reported and actual scores, often an indicator of validity of the self-reported data (Cassady 2001; Kuncel et al. 2005). To examine whether or not ACT self-reported scores had a higher degree of validity compared to SAT self-reported scores, a formula was used that tests differences in correlations for independent samples where the sample sizes are different (Cohen 2001) (Fig. 2). In the equation below, Z_{r_1} and Z_{r_2} are the correlations of the differences to be tested and N_1 and N_2 are the corresponding sample sizes. Since a z score is needed for both the difference in correlations between SAT total/SATCR and SAT total/SATMR, a Bonferroni adjustment was made to guard against a Type I error (Cohen 2001). With this adjustment (.05/3) and that the transformed z critical value that corresponds to the critical values of a t distribution, the critical value must be significant at $p < .015$ ($Z_{cv} = 2.45$).

Fig. 2 Test of differences in correlations for independent samples. $z = \frac{z_{r_1} - z_{r_2}}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}}$ (Cohen 2001, p. 269)

$$z = \frac{Z_{r_1} - Z_{r_2}}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}}$$

(Cohen, 2001, p. 269)

Data Analysis

All three of the indicators of self-reported data accuracy described above were examined for research questions 1 and 2. For research question 1, frequency of error was examined by comparing the frequency of students under-reporting, accurately reporting, and over-reporting their test scores. Mean differences in the magnitude of bias were examined using *t*-tests. Finally, accuracy was investigated by testing the differences in self-reported-to-actual correlations for the ACT as compared with each of the three SAT test sections.

For research question 2, students were grouped using percentile rankings into three achievement levels that represented the lower third (low), the middle third (medium), and upper third (high) of test scores as reported by the institution. Frequency of error was examined by comparing the frequency of students under-reporting, accurately reporting, and over-reporting their test scores by each achievement level. Analysis of variance (ANOVA) was used to determine significant differences in the magnitude of bias for each test score. The dependent variables were the three individual SAT test scores, the SAT Total score, and the ACT Composite score. The factors included the variable indicating whether or not they over- or under-reported and the students' test scores. The primary focus of this analysis was to look for an interaction between reporting accuracy and achievement, indicating that there was systematic bias by achievement level. For this study, the presence of a significant interaction indicated that the magnitude of bias was significantly different between those students who over-report and those who under-report their test scores. Finally, accuracy was investigated by testing the differences in correlations for independent samples.

For research question 3, a paired-samples *t*-test was used to determine if there were significant differences between the bias in the SAT Critical Reading and Mathematics sections and the SAT Total score. Also, accuracy was investigated by testing the differences in self-reported-to-actual correlations between the SAT Total score and scores for the individual sections.

Results

Research Question 1: Is there a difference between the accuracy of self-reported ACT scores and the accuracy of self-reported scores for the three SAT Reasoning Test sections?

Frequency of Error

For each SAT Reasoning Test section (Critical Reading, Mathematics, and Writing), the percentage of students that reported accurate scores (within plus or minus 20 points) ranged from 69.7% for the Mathematics section to 61.5% for the Writing section (Table 2). Also, for each of the three SAT sections, the percentages indicated that students were somewhat more likely to over-report than to under-report, particularly with the Critical Reading and Writing scores. For the Critical Reading scores, approximately 21% of the students over-reported their scores, compared to 15% that under-reported. For the Writing scores, approximately 22% of the students over-reported their scores, compared to 17% that under-reported. There was little difference in the percentage of students that over- or under-reported their Mathematics scores (16% over-reported compared to 15% under-reported). In contrast, 89% of the ACT Composite scores were reported accurately (within

Table 2 Frequency and descriptive statistics associated with under-reporting, accurate reporting, and over-reporting

		Under	Accurate	Over
SAT critical reading	Percent	14.6	64.7	20.7
	M ^a	551.1	560.6	503.9
	SD	85.7	88.3	84.8
SAT mathematics	Percent	14.6	69.7	15.7
	M ¹	548.8	580.7	524.5
	SD	84.3	87.1	88.4
SAT writing	Percent	16.5	61.5	21.9
	M ^a	545.0	559.2	502.1
	SD	85.4	85.6	83.6
ACT composite	Percent	1.6	88.7	9.7
	M ^a	24.6	23.8	21.5
	SD	4.4	4.3	4.1

^a Mean of institution-reported test scores

plus or minus 1 point), but most (10%) of those who reported inaccurately over-reported their scores.

These results support the cognitive distortion explanation for error in self-reported test scores which predicted that ACT scores would be more accurately recalled than any of the SAT scores because the ACT score provides less information to retrieve. These results also provide some support for the motivated distortion view which predicted that when students are not accurate, they tend to overestimate rather than underestimate their scores. This was the case for two of the SAT scores and the ACT score, but not for the SAT Mathematical Reasoning score.

Magnitude of Bias

For those students who under-reported their SAT scores, the average magnitude of bias ratio ranged from .11 to .12 and for those who over-reported their SAT scores, the average bias ratio ranged from .11 to .12. *t*-test comparisons of these values for over- and under-reporters revealed no significant differences in the magnitude of bias for student self-reports of their SAT Critical Reading, Mathematics, and Writing scores (Table 3). The average bias for those who under-reported their ACT score was 14.1%, while for those who over-reported the average bias was 13.6%, and *t*-test results indicated that these values were not significantly different.

Correlation

The final indicator of overall accuracy concerned the validity of the self-reported data, which was investigated by examining differences in the correlations between self-reported and institution-reported scores.

At the aggregate level, scores display generally high correlations for all tests ranging from .86 to .95 (Table 4). However, the correlation between self-reported and actual ACT scores is significantly higher than any of the SAT individual test correlations ($p < .001$). As with the other two indicators of accuracy, these results provide support for the cognitive distortion view in that it appears that students are significantly more accurate in recalling their ACT score than students who try to recall each of their three SAT scores.

Table 3 Magnitude of bias

		Under	Over	<i>t</i> value
SAT critical reading	Mean bias	.112	.114	-.596
	SD	.079	.077	
SAT mathematics	Mean bias	.114	.113	.477
	SD	.084	.078	
SAT writing	Mean bias	.116	.120	-1.340
	SD	.084	.080	
ACT composite	Mean bias	.141	.136	.946
	SD	.104	.082	

Table 4 Correlation between self-reported and actual test scores and if SAT correlations are significantly different than ACT correlation

	Overall <i>r</i>	Difference from ACT
SAT critical reading	.86	.09***
SAT mathematics	.88	.07***
SAT writing	.88	.07***
ACT composite	.95	–

*** *p* < .001

Table 5 Frequency of under, accurate, or over-reporting by achievement level

		Accuracy group			% difference (over–under)
Achievement level		Under (%)	Accurate (%)	Over (%)	
SAT critical reading	Low	13.7	52.3	34.0	20.3
	Medium	13.3	65.5	21.2	7.9
	High	15.9	72.3	11.8	-4.1
SAT mathematics	Low	17.7	57.1	25.2	7.5
	Medium	15.9	66.6	17.5	1.6
	High	12.0	78.9	9.1	-2.9
SAT writing	Low	17.0	45.6	37.5	20.5
	Medium	15.7	61.3	23.1	7.4
	High	17.1	71.5	11.5	-5.6
ACT composite	Low	1.1	83.4	15.5	14.4
	Medium	1.4	88.6	10.0	8.6
	High	1.9	92.2	5.9	4.0

Research Question 2: Is there any difference between the accuracy of self-reported ACT and SAT scores based on the achievement level of the student?

Frequency of Error

Table 5 lists the frequencies of reporting accuracy on the SAT and ACT tests, and can be examined in several ways. First, the frequencies in the “Accurate” column (those within 20 points plus or minus on the SAT and 1 point plus or minus on the ACT) strongly indicate that high-achieving students were more likely to self-report their test scores

accurately. For example, on the SAT Critical Reading section, approximately 52% of students who scored in the lowest third of the distribution were accurate, while more than 72% of those in the highest third were accurate. The other two SAT sections and the ACT Composite showed similar patterns, with the accuracy of low achievers well below that of high achievers on each test.

Looking at the results in Table 5 a different way, low-achieving students were much more likely to overestimate their scores than to underestimate, whereas high-achieving students were more balanced in their occurrences of under- or overestimating. For example, over one-third (34%) of students scoring in the lower third of the SAT Critical Reading section over-reported their test results, while less than half of that number (13.7%) under-reported. In contrast, slightly fewer high-achieving students over-reported (11.8%) than under-reported (15.9%). ACT test takers on the other hand, were more accurate but were also more likely to overestimate their scores across all three achievement levels. Very few ACT test-takers underestimated their scores. These results support for the motivated distortion explanation for errors in self-reported test scores for lower achieving students.

Magnitude of Bias

Table 6 reports the magnitude of bias by achievement level for each SAT section and the ACT, and also reports the tests of interaction between achievement level and accuracy group. On all three SAT sections and the ACT composite, the low-achieving students who over-reported their test scores displayed the greatest amounts of bias in their data. So not only were low achievers less likely to be accurate as described previously, they were inaccurate to a greater degree when self-reporting their SAT scores. Students who under-reported their scores tended to be off by a similar magnitude across the three achievement groups supporting the hypothesis that distortion for these students in more cognitive in nature than motivated (ACT results for those who under reported do not fit this pattern, however so few ACT test takers under reported their scores that generalizing results by achievement level are tenuous). All interactions were significant ($p < .001$) providing support for the motivated distortion hypothesis that low-achieving students systematically over-report their test scores to a greater degree than their higher achieving peers. For example, low-achieving students who over-reported their Critical Reading scores did so with an average bias of .132, whereas high-achieving students who over-reported their scores reported a mean bias of .09. There was little difference in bias for students who under-reported their scores regardless of achievement level.

Correlation

As indicated in Table 7, the correlations between self-reported and institution-reported test scores are significantly lower for low and medium-achieving students compared to high-achieving. Interestingly, correlations for low achievers ranged between .47 and .68 while the medium group correlations were even lower (.35 to .63). Though correlations between self-reported and institutional scores for high-achieving students are generally high and consistent with similar prior research, the self-reported data appears to be much less valid (predictive of actual scores) for the lower two-thirds of test takers.

Research Question 3: Is there any difference between the accuracy of a calculated combined SAT score compared to the individual self-reported SAT scores?

Table 6 Magnitude of bias by achievement level

Test	Accuracy group	Achievement level	Mean bias	SD	Tests of interaction
SAT critical reading	Under	Low	.113	.071	SATCR *achievement level: F = 16.18; $p < .001$
		Medium	.111	.080	
		High	.113	.083	
	Over	Low	.132	.094	
		Medium	.103	.053	
		High	.094	.052	
SAT mathematics	Under	Low	.124	.076	SATMR *achievement level: F = 11.08; $p < .001$
		Medium	.118	.096	
		High	.104	.076	
	Over	Low	.138	.098	
		Medium	.100	.059	
		High	.090	.046	
SAT writing	Under	Low	.121	.074	SATWR *achievement level: F = 7.82; $p < .001$
		Medium	.117	.096	
		High	.112	.078	
	Over	Low	.134	.097	
		Medium	.113	.068	
		High	.098	.049	
ACT composite	Under	Low	.145	.098	ACT *achievement level: F = 18.31; $p < .001$
		Medium	.113	.055	
		High	.157	.124	
	Over	Low	.148	.097	
		Medium	.141	.079	
		High	.108	.045	

Table 7 Correlations between self-reported and actual test scores by achievement level

Achievement level	SAT critical reading	SAT mathematics	SAT writing	ACT composite
Low	.49	.50	.47	.68
Medium	.35	.40	.36	.63
High	.77	.82	.75	.91
Difference (High–Low)	.28***	.32***	.29***	.24***

*** $p < .001$

Magnitude of Bias

As indicated in Table 8, the self-reported SAT Total (sum of self-reported Critical Reading and Mathematics scores) had significantly less bias than for either of the individual SAT test sections for students who both over- and under-reported their scores. The mean bias for the self-reported SATCR and the SATMR ranged between .113 and .116 for both under- and over-reporters, whereas the mean bias for the self-reported SAT Total ranged between .078 and .088 for both under- and over-reporters. These results provide evidence that

Table 8 Mean differences in magnitude of bias between sat total, critical reading, and mathematics

Magnitude of bias				
Accuracy group	SAT critical reading	SAT total	Difference	<i>t</i> -value
Over	.115	.078	.037	29.723***
Under	.113	.079	.034	21.383***
	SAT mathematics	SAT total	Difference	<i>t</i> -value
Over	.114	.088	.026	17.251***
Under	.116	.078	.038	23.948***

*** $p < .001$ **Table 9** Correlation between self-reported and actual test scores and if SATCR and SATMR correlations are significantly different than SAT Total correlation

	Overall <i>r</i>	Difference from SAT Total	<i>t</i> -value
SATCR	.857	.032	11.298***
SATMR	.880	.009	3.404***
SAT total	.889	–	–

*** $p < .001$

random error in self-reported SATCR and SATMR scores are possibly reduced if these scores are combined into one total score. Also, an alternative explanation could be that students may actually know their SAT total score better than the individual subject scores. Thus, students may recall the total score more accurately and then deduce the individual test scores less accurately when asked to do so on the survey. This alternative explanation supports the cognitive distortion idea because they may remember the single SAT score better than the separate test scores.

Correlation

Again, at the aggregate level, SAT scores displayed a high level of correlation (Table 9). However, the correlation between self-reported and institution-reported Critical Reading and Mathematics scores was significantly lower than the correlation between self-reported and institution-reported SAT Total scores ($p < .001$). These results provide support, though minimal, that the combining of self-reported data reduces the amount of error in the measurement.

Limitations

It should be acknowledged that at least some of the bias in inaccuracies may be due to floor and ceiling effects of the possible scores themselves. That is, fewer high-achieving students may be expected to overestimate their scores because fewer scores above their institutional scores were available to select. Likewise, fewer low-achieving students may be expected to underestimate their scores because fewer scores below their institutional scores existed. However, if floor and ceiling effects were to explain all the biases in reporting, it follows logically that high achievers should under-report as often as low

achievers over-report, which is not true. Again, low achievers over-report at much higher rates, while the percent of under-reporting among low, medium, and high achievers is fairly balanced. So while floor and ceiling effects may be a factor, they do not account for the bulk of the phenomenon we observe.

Also, as acknowledged in the Method section, many students take the SAT and ACT tests more than once. For these students there could be more than one accurate test score, and in the present study it was not known which scores was reported to NSSE. Though this study attempted to account for the ‘error’ that may be introduced by students who have more than one test score, there is no way to know to what extent our measurement of accuracy still did not account for scores from multiple test-takers.

Discussion

The purpose of this study was to investigate the accuracy of self-reported SAT and ACT scores overall and by achievement level. This study confirmed prior research which found that when students are inaccurate in reporting their scores, a disproportionate number of them over-report their scores indicating that some students may systematically exaggerate their scores. Also the correlation between self-reported and actual SAT section scores were significantly lower than correlations for self-reported and actual ACT Composite scores indicating that students may experience some cognitive distortion when recalling their SAT scores which contain more complex information to retrieve. As expected the concurrent validity of self-reported test scores was high, as indicated by high correlations between self-reported and actual scores. Also consistent with other studies, this study found that lower achieving students for both tests were much less accurate when reporting their scores, again providing evidence that some students systematically inflate their scores.

There are at least four implications of for higher education research:

1. *In cases where students took both the ACT and SAT tests, researchers should favor self-reported ACT scores over SAT scores.* This study found compelling evidence that students report their ACT scores more accurately than students who report SAT scores.
2. *Researchers should be cautious when creating groups of students based in self-reported test scores.* For instance, a researcher interested in studying low-achieving students may not want to identify these low-achieving students based solely on self-reported test scores. Since these students are much more likely to overestimate their test scores, to identify them based solely on their scores would likely underestimate the relationship of their test scores with other variables included in the study.
3. *Summing the self-reported SAT Critical Reading and Mathematics scores to create a combined SAT score appears to reduce measurement error.* This reduction in measurement error may be due to the cancelling out of random errors that are present in each of the two self-reported scores. Given these results it is recommended that researchers combine SAT scores to create an SAT Total score in order to increase the accuracy of these data.
4. Researchers should exercise caution when using self-reported test scores as covariates to control for prior learning. Multivariate analysis assumes that covariates are measured without (or at least minimal) error (Tabachnick and Fidell 2001). However, while almost all variables contain measurement error, the purpose for including a covariate is to reduce the unexplained variance thereby increasing the sensitivity of the

test of main effects. However, if the covariate itself has excessive measurement error, then the researcher may well be introducing more unexplained error into the model rather than controlling for error. In regression, measurement error in the independent variables can lead to upward or downward estimation of the regression coefficient (Pedhazur 1997). In other words, excessive measurement error can lead to increased probability of a Type II error (not rejecting the null when it is not true). For instance, the use of self-reported test scores to predict first-year GPA for low-achieving students may lead the researcher to conclude that there is no significant relationship between incoming test scores and first-year GPA. When in fact, given the amount of error in self-reported test scores for low-achieving students, test scores may be significantly predictive of first-year GPA.

With this caution noted, this study found that the self-reported ACT Composite score contained relatively little error. Also, this study found that the combined SAT score (sum of the Critical Reading and Math) did in fact reduce the error compared to the self-reported sub-test scores. Given these results, researchers can use self-reported ACT composite scores with confidence. However, with regard to the SAT, researchers may want to request that students report their total score, rather than their individual sub-test scores. Though additional research is needed to confirm, given the results of this study, it seems reasonable to assume that students will report their total combined SAT score more accurately than their individual sub-test scores.

The use of self-reported data in higher education research will likely continue for the foreseeable future. This study provided additional important information regarding the accuracy of these data and their appropriate use in higher education research.

References

- Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Upper Saddle River, NJ: Prentice Hall.
- Anaya, G. (1999). Accuracy of self-reported test scores. *College & University*, *Fall*, 13–19.
- Bahrnick, H. P., Hall, L. K., & Berger, S. A. (1996). Accuracy and distortion in memory for high school grades. *Psychological Science*, *7*, 265–271.
- Beretvas, S., Natasha, Meyers, J. L., & Leite, W. L. (2002). A reliability generalization study of the Marlowe-Crowne social desirability scale. *Educational and Psychological Measurement*, *62*, 570–589.
- Braxton, J. M., Brier, E. M., & Hossler, D. (1988). The influence of student problems on student withdrawal decisions: An autopsy on “autopsy” studies. *Research in Higher Education*, *28*, 241–253.
- Cabrera, A. F., Stampen, J. O., & Hansen, W. L. (1990). Exploring the effects of ability to pay on persistence in college. *The Review of Higher Education*, *13*, 303–336.
- Cassady, J. C. (2001). Self-reported GPA and SAT: A methodological note. *Practical Assessment, Research, & Evaluation*, *7*. Retrieved April 17, 2008, from <http://PAREonline.net/getvn.asp?v=7&n=12>.
- Cohen, B. H. (2001). *Explaining psychological statistics* (2nd ed.). New York, NY: John Wiley & Sons.
- College Board. (2008). *SAT-ACT concordance tables*. Retrieved Aug 15, 2008, from <http://professionals.collegeboard.com/profdownload/act-sat-concordance-tables.pdf>.
- Creswell, J. W. (2008). *Educational research: planning, conducting, and evaluating quantitative and qualitative research* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Dobbins, G. H., Farh, J., & Werbel, J. D. (1993). The influence of self-monitoring on inflation of grade-point averages for research and selection purposes. *Journal of Applied Social Psychology*, *23*, 321–334.
- Frucot, V. G., & Cook, G. L. (1994). Further research on the accuracy of students’ self-reported grade point averages, SAT scores, and course grades. *Perceptual and Motor Skills*, *79*, 743–746.
- Gonyea, R. M. (2005). Self-reported data in institutional research: Review and recommendations. In R. K. Toutkoushian & P. D. Umbach (Eds.), *Survey research: emerging issues: Vol. 127. New directions for institutional research series* (pp. 73–90). San Francisco, CA: Jossey-Bass.

- Gramzow, R. H., & Williard, G. (2006). Exaggerating current and past performance: Motivated self-enhancement versus reconstructive memory. *Personality and Social Psychology Bulletin*, *32*, 1114–1125.
- Herman, W. E. (2003, August). *College student awareness of current G.P.A.* Paper presented at the annual meeting of the American psychological association, Toronto, Ontario.
- Kuncel, N. R., Crede, 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.
- Mayer, R. E., Stull, A. T., Campbell, J., Almeroth, K., Bimber, B., Chun, D., et al. (2007). Overestimation bias in self-reported SAT scores. *Educational Psychology Review*, *19*, 443–454.
- Nancarrow, C., & Brace, I. (2000). Saying the “right thing:” coping with social desirability bias in marketing research. *Bristol business school teaching and research review*, Summer, 3.
- Pearson, R. W., Ross, M., & Dawes, R. M. (1994). Personal recall and the limits of retrospective questions in surveys. In J. M. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys* (pp. 65–94). New York, NY: Russell Sage Foundation.
- Schmidt, F. L., Le, H., & Ilies, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual differences constructs. *Psychological Methods*, *8*, 206–224.
- Shepperd, J. A. (1993). Student derogation of the scholastic aptitude test: Biases in perceptions and presentations of College Board scores. *Basic and Applied Social Psychology*, *14*, 455–473.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Boston, MA: Allyn & Bacon.
- Takalkar, P., Waugh, G., & Micceri, T. (1993, May). *A search for truth in student responses to selected survey items*. Paper presented at the AIR annual forum, Chicago, IL.
- Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response*. Cambridge; New York: Cambridge University Press.
- Williard, G., & Gramzow, R. H. (2008). Exaggeration in memory: Systematic distortion of self-evaluative information under reduced accessibility. *Journal of Experimental Social Psychology*, *44*, 246–259.