The Accuracy of Self-Reported Achievement Scores among High School Students in Addis Ababa, Ethiopia

Seleshi Zeleke[1], Mekonnen Guale[2] and Mitiku Hambisa[3]

*DOI: DOI: https://doi.org/*10.63990/ejobs.14304

Received: 27 July 2026; Accepted: 08 September 2026

Abstract

Researchers in various disciplines, including psychology and education, usually use self-report measures to collect data from research participants. They prefer to use self-report data because they are easily accessible and more efficient. This study examined the accuracy of high school students' self-reported mathematics, English, and chemistry achievement scores from correlational and discrepancy perspectives. Data were collected from 123 twelfth-grade students drawn from two (one public and one private) high schools in Addis Ababa. Self-reported achievement scores were collected from the students using a questionnaire whereas the corresponding scores were collected from official school records. The results indicated that the correlations between self-reported scores and the corresponding actual scores range from .59 to .81. On the other hand, the discrepancy perspective yielded statistically significant mean differences for all pairs of self-reported and actual scores. The self-reported mean scores were significantly larger than the corresponding actual mean scores. A significantly larger proportion of the students over-reported their scores than those who underreported or accurately reported their scores. Over-reporting of scores was more common among low achievers than among average or high achievers. The number of students who accurately reported their mid-semester scores was larger than the number of students who accurately reported their first or second semester scores. The overall pattern was similar across the three school subjects. In conclusion, whereas self-reported scores were more or less valid from the correlational perspective, their accuracy from the discrepancy perspective is concerning.

Keywords: self-reported achievement scores, actual scores, accuracy, validity, discrepancy

[1] Corresponding author, Associate Professor, School of Psychology, CELS, Addis Ababa University. Email: sileshi.zeleke@aau.edu.et

[2] Doctoral candidate, Department of Educational Planning and Management, CELS, Addis Ababa University

[3] Assistant Professor, School of Psychology, CELS, Addis Ababa University. Email: lovemitiku@yahoo.comt

Background to the Problem

Researchers in various fields, including psychology and education, usually use self-report measures to collect data from research participants. Self-report measures are commonly used to collect information that is not available in databases or in the participants' personal files that are officially recorded. Even when databases or official records are available, researchers often prefer self-reported data because of their accessibility and efficiency (Sticca et al., 2017).

Whereas many of the questions in self-report measures are attitudinal in nature, some ask participants to answer factual questions. For example, self-report measures may ask students their attitudes toward learning mathematics. Factual questions, by contrast, may ask students about the actual scores they earned in mathematics in the first semester. Factual and attitudinal questions are distinct in that the former have correct answers while the latter do not (Rosen et al., 2017).

In general, we cannot verify the accuracy of participants' responses to attitudinal questions. However, we can verify responses to factual questions because they have correct answers. For example, we can verify the accuracy of students' responses about the scores they earned in school subjects like English and mathematics by looking at their actual scores in official school records (which are presumed to be accurate).

Because of this possibility of verification of the accuracy of self-reported academic achievement scores, researchers raise an important question. To what extent do students accurately report their achievement scores when asked for research purposes? A growing number of studies predominantly from North America have examined the accuracy of self-reported data pertaining to high school and college students' GPA, SAT Scores, and high school GPA (e.g., Cassady, 2001; Cole & Gonyea, 2010; Rosen et al., 2017) as the literature review below shows.

The literature commonly uses two perspectives to examine the accuracy of self-reported data. Sticca et al. (2017) identify these as correlational and discrepancy perspectives. The correlational perspective is concerned with the extent to which actual data (e.g., data from a record) and self-reported data are correlated. Consistent with Classical True Score Theory, recorded data and self-reported data can be treated as true and observed scores, respectively, and estimates of the correlations between these two sets of data can be used to estimate the reliability of the self-reported data (Kuncel et al., 2005). Accordingly, several studies have examined the

link between students' self-reported and actual grades and found strong, positive, statistically significant results.

For instance, Cassady (2001) found students' self-reported GPA remarkably similar to those from official records; the correlation between self-reported and actual cumulative GPA was $r = .97$ and statistically significant ($p < .001$). More specifically, Cassady found strong and positive relationships between self-reports and actual verbal performance subscale ($r = .73$, $p = .001$) and math subscale ($r = .89$, $p = .001$). Escribano and Díaz-Morales (2014) studied 1,121 secondary school adolescents in Madrid (Spain) and found statistically significant positive associations between self-reported and actual measures; the coefficients were similar for girls and boys (math: $r = .66$ vs. $r = .58$; English: $r = .70$ vs. $r = .73$; social sciences: $r = .60$ vs. $r = .68$ and Spanish: $r = .64$ vs. $r = .64$). Likewise, Cole and Gonyea (2010) found statistically significant positive correlations between self-reported and actual test scores. These results appear to corroborate Kuncel et al.'s (2005) earlier findings.

In a meta-analytic study of 60,926 participants from 37 independent high school and college level student samples, Kuncel et al. (2005) found .84 to be the average correlation between actual grades and self-reported scores; with correlations for college students ($r = .90$; $N = 12,089$) being greater than that of high-school students ($r = .82$; $N = 44,176$).

Nonetheless, some critiques (e.g., Rosen et al., 2017) argue that many scholars erroneously conclude that student self-reports are accurate due to an overreliance on bivariate correlations. As an example of their argument, Rosen et al. used Kuncel et al.'s (2005) finding that the correlation between actual and self-reported GPA is .82. This relationship sounds robust, argued Rosen et al., until one estimates the percentage of variance in self-reports due to actual GPA. If students' self-reported GPA is regressed on actual GPA, an r of .82 indicates that the R^2 from this regression model would be .67. This implies that, Rosen et al. continued, actual GPA explains only two-thirds of the variance in self-reported GPA. This, in turn, implies that self-reports contain substantial error. If this were random error, we would be less concerned, but the literature is clear that this is not the case, as there are substantial patterns of over-reporting among lower-performing students. The discrepancy perspective addresses this issue of over-reporting and underreporting.

From the discrepancy perspective, Sticca et al. (2017) contend that the accuracy of a student's self-reported grade is represented by the difference between self-reported and actual grades (i.e., self-reported minus actual grades: *sg-ag*). For this measure (i.e., *sg-ag*), a negative value reflects under-reporting, a positive value reflects over-reporting, and zero reflects an accurate report. A plethora of studies have examined accuracy of self-reported scores using the discrepancy perspective. Indeed, as early as 1993, Dobbins et al. revealed that students with lower GPAs tended to inflate their scores more than students with higher GPAs did. Similarly, Maxwell and Lopus (1994) found that students with below-average grades were most likely to misreport their scores. Other authors have replicated Dobbins et al. (1993) and Maxwell and Lopus's (1994) results over the years (see Rosen et al., 2017).

For instance, Cassady (2001) studied a sample of undergraduate students and found that compared to students in higher-performance categories, the lowest-performing students were much less accurate reporters. Likewise, Zimmerman et al. (2002) reported widespread self-reporting inaccuracy and a tendency by lower-performing students to over-report GPA. Mayer et al. (2007) further found systematic over-reporting in SAT scores, particularly among lower-scoring students. Moreover, Cole and Gonyea (2010) investigated reporting accuracy in self-reported ACT scores and found that when students report inaccurately, a disproportionate number over-report their scores and that lower-achieving students are much less accurate when reporting their scores.

Specifically examining proportions, Escribano and Díaz-Morales (2014) found that 36.4% of students reported their grades accurately, 31.1% of students over-reported their grades whereas 17.2% of the students under-reported their grades. Besides, Escribano and Díaz-Morales's study showed that self-reported grades were more reliable indicators of actual grades for students with higher levels of academic performance and cognitive ability. Moreover, in their longitudinal study of accuracy of self-reported academic grades in mathematics, German, English, and French in a sample of Swiss high school students from grades 9 through 11, Sticca et al. (2017) found a consistent tendency for students to over-report their grades over time and across academic subjects, a finding that they claimed to support and extend previous research.

Notwithstanding the issue of misreporting among lower-performing students, few studies on the use of self-reported measures of academic performance offered some hope in that they demonstrate that self-reported academic achievement scores show acceptable accuracy (Rosen et

al., 2017). From both correlational and discrepancy perspectives, Sticca et al. (2017) demonstrated the accuracy of self-reported grades as indicators of actual grades in the Swiss school system.

The above review clearly shows that, with the exception of a few studies that were conducted in Europe, most were from North America. Studies conducted in Africa (e.g., Nkhoma, 2025) in general and Ethiopia (e.g., Alebie et al., 2026; Fuller et al., 2023) in particular investigated the accuracy of self-reported data in the area of health and development. In general, the studies found discrepancies (i) between self-reported data reported about the same thing at different times and (ii) self-reported data and examination results or clinicians' professional opinions. The authors emphasized the need for caution in interpreting the findings and drawing conclusions from the data.

The reasons behind misreporting appear to be less substantiated in the empirical literature. Researchers suggest different factors to explain misreporting. For example, Sticca et al. (2017) indicate that the tendency of over-reporting is related to self-enhancement or impression management strategies that may enhance positive mood, self-concept, and self-esteem and that this tendency may also put students at risk for developing maladaptive learning strategies and poor performance. Other researchers indicate factors including imperfect storage and retrieval from memory (Kuncel et al., 2005), carelessness when responding (Rosen et al., 2017), and social desirability bias (Cassady, 2001) as explanations for misreporting.

Overall, despite the growing reliance on respondents' self-reports, very little is known about the accuracy of self-reported achievement scores in different contexts. Particularly in the Ethiopian context, it appears that no study has investigated the accuracy of self-reported achievement scores even though few studies have addressed the accuracy of self-reported data in the area of health. Besides, we know very little about whether the accuracy of self-reported achievement scores varies across school subjects. Furthermore, we do not know much about whether some variables (e.g., gender, achievement level) moderate the accuracy of self-reported scores. The purpose of this study is therefore to examine the accuracy of self-reported achievement scores in English, mathematics, and chemistry among twelfth-grade students in Addis Ababa, Ethiopia in an attempt to narrow the mentioned gaps.

More specifically, this study sought to answer the following research questions.

1. How valid (or accurate) are self-reported achievement scores?

2. Are there statistically significant differences between self-reported scores and the corresponding scores collected from school records for the three school subjects (i.e., English, math, and chemistry)?
3. Are there identifiable patterns of over-reporting, accurate reporting, and underreporting of scores among the study participants?
4. Are there statistically significant differences in accuracy of reporting due to school type and level of academic achievement?

Answers to these questions are expected to shed light on how high school students in the Ethiopian context report their achievement scores when asked for research purposes. That is, the study would clarify the extent to which students underreport, over-report or accurately report their scores. This, in turn, has important implications for research as most research in education, psychology and the social sciences largely depends on self-reported data. The findings of this study can thus provide information regarding how confident researchers could be when doing and reporting research with self-reported data. In general, the findings would shed light on how dependable self-reported data are in general and to what extent researchers need to be cautious when drawing conclusions and implications from research that employed self-reported data.

Conceptual Framework

The two sets of scores (i.e., self-reported scores and the corresponding scores from official school records) in English, math, and chemistry represent the students' academic performance or achievement scores in the three school subjects. They therefore represent the same set of variables in two versions (i.e., self-report and document versions). For this reason, correlations between the two sets of scores yield estimates of reliability of self-reported scores according to Classical Test Theory (Cassady, 2001; Kuncel et al., 2005).

From another perspective, self-reported scores and the corresponding scores from official school records carry different measurement mechanisms and sources of error. While scores from official school records are minimally affected, self-reported scores are likely to be affected by several factors. These include imperfect storage and retrieval from memory (Kuncel et al., 2005) or errors in memory (Cassady, 2001), carelessness on the part of respondents when responding (Rosen et al., 2017), recency or the time elapsed between storing the information and trying to retrieve it to report for research purposes (Kuncel et al., 2005) and self-enhancement needs and

social desirability bias (Cassady, 2001), which lead students to present themselves favorably. Because of the above sources of error, self-reported scores comprise a larger amount of error than scores from official school records (Kuncel et al., 2005; Shaw & Mattern, 2009)

Because of the amount of error they contain, the literature considers the two sets of scores as different variables according to Generalizability Theory. From this perspective, the correlations of the two sets of scores represent validity rather than reliability. On the other hand, the discrepancy or mean difference between the two sets of scores reflects the direction and magnitude of error associated with self-reported scores. Following the literature in the area (e.g., Hatos & Gyarmati, 2023), this study treats correlations between the two sets of scores as validity of self-reported scores and discrepancy of the scores as the magnitude of error.

Finally, based on the amount of error involved in the two sets of scores, previous studies commonly consider the scores from official school records as the benchmark against which the accuracy of self-reported scores is measured (Kuncel et al., 2005; Shaw & Mattern, 2009). Thus, associated with this, if the mean of self-reported scores is higher than the mean of scores from official school records, it is termed over-reporting in this study. If, on the other hand, the self-reported mean score is lower than the mean score from official school records, this is underreporting. Finally, if the two means are equal or the difference between the two means is equal to zero, this is termed accurate reporting. This use of terminology is consistent with the literature in the area (e.g., Kuncel et al., 2005; Sticca et al., 2017; Zimmerman et al., 2002).

Method

Design of the Study

This study employed a correlational design. More specifically, we examined the accuracy of self-reported achievement scores from two perspectives: correlational and discrepancy perspectives. The variables we examined included (1) self-reported achievement scores in three school subjects (that is, English, mathematics, and chemistry), (2) the corresponding scores from official school records, (3) achievement level, and (4) school type.

Participants

First, the researchers selected two (one government and one private) high schools from which they could easily collect data. That is, in addition to collecting data from the students directly (mainly self-reported achievement scores), the researchers had to collect the corresponding scores from official school records. Because of the administrative restrictions in many schools on accessing students' achievement scores from official school records, the researchers had to select the two high schools purposefully. That is, they selected the two high schools where it was more convenient for them to collect the necessary data. More specifically, one of the researchers was acquainted with the directors of the two schools.

Second, to avoid dealing with students from different grades (and to avoid introducing a variable due to grade differences), the researchers selected Grade 12 students only (in the 2021/22 Academic Year). Third, focusing on the natural science stream students only (i.e., three sections in the private high school and five sections in the government high school), we selected one section from each school randomly using a lottery method. We also studied natural science stream students to avoid introducing still another variable due to possible differences between natural and social science stream students.

Fourth, all students from the two randomly selected sections who were present on the date of data collection participated in the study. Thus, we collected data from 125 (59 female and 66 male) students. However, because two participants provided incomplete data for many of the variables, we excluded them from analysis. Thus, the analysis was based on data from 123 (59 female and 64 male) students. The age of the students ranged from 17 to 20 (M = 18.02 years, SD = 0.62).

Tools for Data Collection

We used a self-administered questionnaire in which the respondents reported their achievement scores for three school subjects (English, mathematics, and chemistry). We asked the students to recall and report three scores for each school subject: (1) their first semester score of the previous academic year (that is, 2020/21), (2) their second semester score of the 2020/21 academic year, and (3) their mid-semester score for the 2021/22 academic year. Thus, we asked each student for three scores per school subject, yielding nine scores for the three school

subjects. After collecting the self-reported scores, we obtained the corresponding nine scores for each student from official school records (henceforth, actual scores).

Data Collection Procedure

In the present study, we randomly selected three school subjects (English, math, and chemistry) for examination using a lottery method. At the time of data collection, nearly a year had elapsed since the students received their first semester scores, whereas nearly six months had elapsed since they received their second semester scores. On the other hand, only two to four weeks had elapsed since the students received their mid-semester scores. In other words, the mid-semester scores were the most recent achievement scores students were asked to report.

We collected the self-reported scores using a self-administered questionnaire. The questions asked the participants to provide their sex, age, and name. In addition, as indicated above, the students were asked to provide their scores in three school subjects: English, mathematics, and chemistry. The names of the participants were needed to collect their actual achievement scores from official school records (roster).

D*ata Analysis Techniques*

We used Pearson r, paired-samples t test, independent samples t test, chi-square test, and quartiles to analyze the data. We used Pearson r to examine the correlation between self-reported and actual scores. We examined discrepancies between the two sets of scores (self-reported and actual scores) using a paired-samples t-test.

In the Ethiopian school system, the maximum possible score in each semester for each school subject is 100 irrespective of the school type. The maximum possible score on the mid-semester test, on the other hand, was 30 for mathematics and chemistry for students from both the government and private schools. However, there was a discrepancy in the mid-semester English score between the government and the private school. The weight allotted for the English mid-semester test was 20% (that is, the maximum possible score was 20) in the private school, whereas the corresponding maximum score was 30 for students in the government school. Thus, to have the scores on the same scale, we converted the scores for the students from the private school to a maximum possible score of 30, which makes the scores of the students from the two schools comparable.

To examine whether accuracy of reporting differs due to the students' level of achievement, we used quartiles to divide the students into four groups in terms of their level of achievement on the actual scores. Then we started the analysis using one-way ANOVA with accuracy of reporting as the dependent variable and level of achievement (i.e., quartiles) as the independent variable. However, because the data violate the homogeneity of variance assumption, we used Welch's ANOVA instead, as it is a robust alternative. As a follow-up to all statistically significant differences in accuracy of reporting, we used Games-Howell's post-hoc test to determine which specific achievement level is more or less accurate in their self-reports. We also used an independent samples t-test to examine mean differences in accuracy of reporting (i.e., deviation scores) between students from the private and government high schools. Furthermore, we used a chi-square test to compare proportions of students.

In addition to the primary objective of the study (which is examining the accuracy of self-reported scores employing correlational and discrepancy perspectives), we designed the study to investigate whether the length of time elapsed between receiving the scores and reporting them for this study makes any difference in accuracy of reporting. If students inaccurately report (i.e., over-report or underreport) their scores because of the length of time elapsed, then possibly they can accurately report their mid-semester scores more so than their first- and second-semester scores because the mid-semester scores were the most recent scores. Similarly, the students can accurately report their second-semester scores more so than their first-semester scores.

Ethical Considerations

We first obtained a support letter from the School of Psychology, Addis Ababa University, describing the purpose of the study and asking the school principals to support them in data collection. Upon presenting the support letter to the principals, the principals gave them permission to collect the necessary data. Accordingly, they collected the self-reported data face-to-face and later obtained the actual data from official school records with the support of the record officers of the two schools.

On the fixed date, the authors met the selected students in a classroom and collected the self-reported data after informing the students of the purpose of the study. The authors also told the students that (i) participation is voluntary and that they can withdraw from the study without penalty even after they started filling in the questionnaire, (ii) the data would be kept

confidential, and (iii) the data would be used for research purposes only. After that, the students were asked to indicate their willingness, and everyone agreed to complete the questionnaire.

Because of the nature of the research, the students were asked to write their names on the self-administered questionnaire so that their actual scores could be easily traced from official school records. Nonetheless, the researchers exerted every effort to keep their names confidential until completing the data entry. After the data entry, the researchers deleted names of the students and the data became anonymized.

Results

Among the 125 students drawn randomly from the two schools for the study, 123 remained after excluding two with extensive missing values, a response rate of 98.4%. However, there were additional item-level missing values, which were handled through pairwise deletion.

Before the actual data analysis using parametric tests, the authors subjected the data to several tests to examine whether the statistical assumptions (namely, normality, homogeneity of variance, independence of observations and absence of outliers) were tenable. The examination confirmed that with the exception of the homogeneity of variance, all the assumptions were tenable. We therefore used Welch's ANOVA, which is a robust test when group variances are not equal. We also used Games-Howell's post-hoc test.

Table 1 presents descriptive statistics for the self-reported scores and the corresponding actual scores. The data show that self-reported scores are consistently higher than the actual scores. Besides, with the exception of the mid-semester scores for chemistry, the actual scores are characterized by more variability than the self-reported scores.

Table 1

Descriptive statistics of the self-reported and actual achievement scores

Score (Semester)	Self-reported score			Actual score		
	N	Mean	SD	N	Mean	SD
English (First)	119	76.32	11.45	120	70.60	14.00
English (Second)	120	78.28	11.36	120	73.98	14.71
English (Mid)	118	19.61	3.77	119	18.21	4.44
Mathematics (First)	119	70.45	11.95	120	63.96	15.28
Mathematics (Second)	120	73.50	11.23	120	70.19	14.17
Mathematics (Mid)	119	16.66	3.89	120	14.73	5.09
Chemistry (First)	119	68.92	10.88	120	60.08	12.10
Chemistry (Second)	120	72.48	11.28	120	66.23	14.85
Chemistry (Mid)	120	16.16	4.73	120	14.61	4.44

Note. There are missing values that range from 3 to 5; that is why the sum in each row is not equal to 123.

Validity of self-reported scores

In the present study, validity of the self-reported scores was estimated by Pearson r, which shows the correlation between self-reported scores and the corresponding scores from official school records. Table 2 presents the correlation between self-reported and actual scores for the first semester, second semester, and mid-semester for each school subject. The data show moderate to strong positive and statistically significant correlations between self-reported and actual scores, indicating the validity of self-reported scores for each of the three school subjects.

A closer look at the validity coefficients in Table 2 indicates that, except for chemistry, the coefficients for the mid-semester test scores (.59 - .67) are lower than those for the first semester (.66 - .77) and second semester (.71 - .81) scores. Besides, for all three subjects, the validity coefficients for the second semester scores are larger than those for the first semester scores.

Table 2

Intercorrelations of self-reported and actual scores and paired-samples t-test results

Score (Semester)	r	T	df
English (First)	.77***	7.25***	116
English (Second)	.81***	5.57***	117
English (Mid)	.59***	3.84***	117
Mathematics (First)	.74***	7.48***	116
Mathematics (Second)	.80***	4.82***	117
Mathematics (Mid)	.65***	5.68***	118
Chemistry (First)	.66***	10.26***	116
Chemistry (Second)	.71***	6.64***	117
Chemistry (Mid)	.67***	4.57***	119

****p < .001*

In addition to the correlational perspective, accuracy of self-reported data can be examined from a discrepancy perspective. Accordingly, we examined the discrepancy between the two sets of scores using a paired-samples t test. The means and standard deviations for both self-reported and actual scores are presented in Table 1. The data in Table 1 clearly show that the self-reported scores are consistently larger than the corresponding actual scores for all nine sets of scores. Besides, examination of the differences of these scores using a paired-samples t test, as shown in Table 2, shows that all the mean differences are statistically significant ($p < .001$) for the three subjects (English, math and chemistry) across the three sets of scores (for the first, second and mid semesters). Overall, the data show that self-reported mean scores are larger with less variability than the corresponding actual mean scores across all nine sets.

Patterns of Over-reporting, Accurate Reporting and Underreporting

In line with Sticca et al.'s (2017) study and from the discrepancy perspective, accuracy of a student's self-reported score in this study is represented by the difference between self-reported and actual scores (i.e., self-reported score minus actual score). Accordingly, negative value reflects underreporting, a positive value reflects over-reporting and zero reflects an accurate

report. In general, the data in Table 3 indicate that most of the students over-reported their scores in the three subjects (English, math and chemistry). The chi-square tests show that the proportion of students who over-reported their scores is significantly larger than the proportion of students who underreported for all nine scores. With the exception of the scores on the mid-semester tests, the proportion of students who over-reported their scores is also larger than the proportion of students who accurately reported their scores. Considering only those who accurately reported their scores, the data show that a significantly larger number of students accurately reported their scores on the mid semester tests than they did with their scores for both the first and second semesters.

Table 3

Frequency and Percentage of Students who Over-reported, Accurately Reported and Underreported their Scores

Score (Semester)	Over-reporting N (%)	Underreporting N (%)	Accurate Reporting N (%)	χ^2
English (First)	78 (66.67)	21 (17.95)	18 (15.38)	58.62***
English (Second)	73 (61.86)	28 (23.73)	17 (14.41)	44.76***
English (Mid)	51 (43.22)	21 (17.80)	46 (38.98)	13.14***
Math (First)	85 (72.65)	21 (17.95)	11 (9.40)	82.67***
Math (Second)	73 (60.33)	32 (26.44)	16 (13.22)	42.86***
Math (Mid)	64 (53.78)	20 (16.81)	35 (29.41)	25.23***
Chemistry (First)	94 (80.34)	8 (6.84)	15 (12.82)	116.97***
Chemistry (Second)	78 (66.10)	24 (20.34)	16 (13.56)	57.83***
Chemistry (Mid)	52 (43.33)	17 (14.17)	51 (42.50)	19.85***

Note. Missing values range from 2 to 6; therefore, the sum in each row is not equal to 123.

***$p < .001$

Variations in Accuracy of Reporting due to School Type

In order to examine accuracy of reporting due to school type, differences in deviation scores (i.e., self-reported scores minus actual scores) were used as the dependent variable, while school type was used as the independent variable. Then, an independent-samples t-test was used to examine the differences. Table 4 summarizes the results obtained for differences due to school type. In general, as indicated in Table 4, the accuracy of reporting among private school students

tended to be significantly better than that of the government school students. More specifically, the accuracy of reporting of the students from the private school was significantly better than that of students from the government school on seven of the nine sets of scores and this is true across the three school subjects.

Table 4
Summary of Independent Samples t-test Results for School Type (Dependent variable=Deviation Score)

Deviation Score	School Type	N	Mean	t	df	Cohen's d
English First Semester	Private	51	1.29	-5.63***	115	0.93
	Government	66	9.65			
English Second Semester	Private	52	1.00	-4.09***	116	0.71
	Government	66	7.23			
English Mid-Semester	Private	53	1.08	-.17	116	-
	Government	65	1.19			
Math First Semester	Private	51	3.73	-3.18**	115	0.57
	Government	66	9.47			
Math Second Semester	Private	52	1.90	-2.17**	116	0.40
	Government	66	5.30			
Math Mid-Semester	Private	54	1.04	-2.59*	117	0.47
	Government	65	2.84			
Chemistry First Semester	Private	51	7.86	-1.25	115	-
	Government	66	10.11			
Chemistry Second Semester	Private	52	2.37	-3.95***	116	0.69
	Government	66	9.61			
Chemistry Mid-Semester	Private	54	.35	-3.33**	118	0.59
	Government	66	2.53			

*p < .05; **p < .01; ***p < .001

Variations in Accuracy of Reporting due to Level of Academic Achievement

Variations in accuracy of reporting were examined by considering the difference in deviation scores (i.e., self-reported scores minus actual scores) as the dependent variable while level of academic achievement (as defined by category of achievement scores in the First, Second, Third and Fourth Quartiles) was used as the independent variable. Then Welch's ANOVA was used to examine the differences. Tables 5, 6 and 7 below summarize the results respectively for English, math and chemistry subjects.

As can be observed from Table 5, statistically significant differences were found in the accuracy of reporting with respect to achievement category in the first (Welch's F = 16.04,

p<.001), second (Welch's F = 16.38, p<.001), and mid (Welch's F = 8.68, p<.001) semester scores for English subject. Follow-up analysis using Games-Howell's post-hoc comparison test indicated that students whose scores fall in the First Quartile (M = 13) significantly over-reported their scores more so than those students in the Third (M = 2.81) and Fourth (M = -.07) Quartiles for English scores in the first semester. Similarly, students in the Second Quartile (M = 8.50) significantly over-reported their scores more so than those students in the Third (M = 2.81) and Fourth (M = -.07) Quartiles for English scores in the first semester.

For English scores in the second semester, Games-Howell's test showed that students whose scores fall in the First Quartile (M = 11.84) significantly over-reported their scores as compared to those students in the Second (M = 5.11), Third (M = 1.68), and Fourth (M = -.97) Quartiles. Likewise, students in the Second Quartile (M = 5.11) significantly over-reported their scores more so than students in the Fourth (M = -.97) Quartile. The results showed similar patterns for English scores in the mid-semester. That is, Games-Howell's post-hoc test showed that students whose scores fall in the First Quartile (M = 3.84) significantly over-reported their scores as compared to students in the Second (M = .90), Third (M = .13) and Fourth (M = -.50) Quartiles. In general, the results for English scores in the first, second and mid semesters indicate that accuracy of reporting increases as the students' level of academic achievement increases.

Tables 5

Summary of Results Obtained from Welch's ANOVA for English Scores

Deviation Score	Achievement Category	N	Mean	Welch's F	η^2
English First Semester (Self-Reported – Actual)	First Quartile	29	13.00	16.04^{***}	0.32
	Second Quartile	28	8.50		
	Third Quartile	32	2.81		
	Fourth Quartile	28	-.07		
English Second Semester (Self-Reported – Actual)	First Quartile	31	11.84	16.38^{***}	0.31
	Second Quartile	27	5.11		
	Third Quartile	31	1.68		
	Fourth Quartile	29	-.97		
English Mid-Semester (Self-Reported – Actual)	First Quartile	31	3.84	8.68^{***}	0.25
	Second Quartile	31	.90		
	Third Quartile	24	.13		
	Fourth Quartile	32	-.50		

$^{***}p < .001$

In a similar manner, Table 6 indicates statistically significant differences in the accuracy of reporting with respect to achievement category in the first (Welch's F = 23.53, p<.001), second (Welch's F = 17.05, p<.001), and mid (Welch's F = 19.74, p<.001) semesters for math subject. Follow-up analysis using Games-Howell's post-hoc test indicated that students whose scores fall in the First Quartile (Mean = 16.28) significantly over-reported their scores more so than students in the second (Mean = 8.60), Third (Mean = 2.87), and Fourth (Mean = -.04) Quartiles for math scores in the first semester.

Table 6

Summary of Results Obtained from Welch's ANOVA for Math Scores

Deviation Score	Achievement Category	N	Mean	Welch's F	η^2
Math First Semester (Self-Reported – Actual)	First Quartile	29	16.28	23.53^{***}	0.38
	Second Quartile	30	8.60		
	Third Quartile	30	2.87		
	Fourth Quartile	28	-.04		
Math Second Semester (Self-Reported – Actual)	First Quartile	30	10.50	17.05^{***}	0.32
	Second Quartile	32	5.63		
	Third Quartile	28	.64		
	Fourth Quartile	28	-2.29		
Math Mid-Semester (Self-Reported – Actual)	First Quartile	30	6.13	19.74^{***}	0.43
	Second Quartile	26	1.58		
	Third Quartile	34	1.12		
	Fourth Quartile	29	-.76		

$^{***}p < .001$

As shown in Table 6, for math scores in the second semester, Games-Howell's post-hoc test showed that students whose scores fall in the First Quartile (Mean = 10.50) significantly over-reported their scores as compared to students in the second (Mean = 5.63), Third (Mean = .64), and Fourth (Mean = -2.29) Quartiles. Similarly, Games-Howell's post-hoc test revealed that students in the Second Quartile (Mean = 5.63) over-reported their second-semester math scores as compared to students in the Third (Mean = .64) and Fourth (Mean = -2.29) Quartiles. The results showed similar patterns for the mid-semester math scores. That is, students whose scores fall in the First Quartile (Mean = 6.13) significantly over-reported their scores more so than students in the Second (Mean = 1.58), Third (Mean = 1.12), and Fourth (Mean = -0.76)

Quartiles. Here also, the results obtained for the first, second, and mid-semester math scores indicate that accuracy of reporting increases as the students' level of academic achievement increases.

As can be observed from Table 7, statistically significant differences were found in the accuracy of reporting with respect to achievement category in the first (Welch's F = 17.39, p<.001), second (Welch's F = 32.01, p<.001), and mid (Welch's F = 4.37, p<.01) semester chemistry scores. Follow-up analysis using Games-Howell's post-hoc test indicated that students in the First Quartile (Mean = 16.34) significantly over-reported their first-semester chemistry scores as compared to students in the Third (Mean = 4.97) and Fourth (Mean = 3.07) Quartiles. Similarly, Games-Howell's post-hoc test showed that students in the Second Quartile (Mean = 11.56) over-reported their first-semester chemistry scores as compared to students in the Third (Mean = 4.97) and Fourth (Mean = 3.07) Quartiles.

Table 7
Summary of Results Obtained from Welch's ANOVA for Chemistry Scores

Deviation Score	Achievement Category	N	Mean	Welch's F	η^2
Chemistry First Semester (Self-Reported – Actual)	First Quartile	32	16.34	17.39^{***}	0.32
	Second Quartile	27	11.56		
	Third Quartile	29	4.97		
	Fourth Quartile	29	3.07		
Chemistry Second Semester (Self-Reported – Actual)	First Quartile	28	18.82	32.01^{***}	0.52
	Second Quartile	31	6.26		
	Third Quartile	28	3.29		
	Fourth Quartile	31	-1.81		
Chemistry Mid-Semester (Self-Reported – Actual)	First Quartile	38	3.50	4.37^{**}	0.14
	Second Quartile	27	.96		
	Third Quartile	28	.93		
	Fourth Quartile	27	.04		

$^{**}p < .01$ $^{***}p < .001$

For Chemistry scores in the second semester, Games Howell's post-hoc test showed that students whose scores fall in the First Quartile (Mean = 18.82) significantly over-reported their scores more so than students in the Second (Mean = 6.26), Third (Mean = 3.29), and Fourth

(Mean = -1.81) Quartiles. Likewise, students in the Second Quartile (Mean = 6.26) significantly over-reported their scores when compared to students in the Fourth (Mean = -1.81) Quartile; also, students in the Third Quartile (Mean = 3.29) significantly over-reported as compared to students in the Fourth (Mean = -1.81) Quartile. Similarly, for mid-semester chemistry scores, Games Howell's post-hoc test showed that students whose scores fall in the First Quartile (Mean = 3.50) significantly over-reported their scores more so than students in the Second (Mean = .96), Third (Mean = .93), and Fourth (Mean = .04) Quartiles. On the whole, the results pertaining to the first, second, and mid-semester chemistry scores indicate that accuracy of reporting increases as the level of academic achievement increases.

Finally, it should be clear that the results are the same across the three school subjects: English, math and chemistry. Thus, for all three school subjects, the accuracy of reporting is significantly better among high achievers than low achievers.

Discussion

This section discusses the main findings of the study in terms of the research questions.

Validity of self-reported scores from the correlational perspective

The validity coefficients of the self-reported scores for the three school subjects range from .59 for English mid-semester scores to .81 for English second–semester scores. Five of the nine coefficients are above .70, and four are below .70. Three of the latter four coefficients pertain to the three mid-semester scores. Overall, the validity coefficients reported in this study are somewhat similar to those of past studies (e.g., Cassady, 2001; Cole & Gonyea, 2010; Escribano & Díaz-Morales, 2014).

In terms of the correlation coefficients, readers should note that the mid-semester scores have generally lower validity coefficients than either the first or second semester scores. To be exact, with the exception of the first semester chemistry scores, the validity of the mid-semester scores are consistently lower than the validity of both the first semester and second semester scores for all three school subjects.

This may be because the maximum possible score for the mid-semester test is 30, while those for the first and second semester scores are 100. This difference would result in restricted range (or limited variability) in the mid-semester scores as compared to the first semester and

second semester scores. In other words, there would be less variability in the mid-semester scores as compared to the variability of the first and second semester scores. Variability is an important factor that affects correlation coefficients. Scores with less variability tend to have lower correlation coefficients as compared to scores with more variability (Mehrens & Lehmann, 1991).

From the correlational perspective, it appears that self-reported scores are generally valid. The mid-semester scores were actually less valid than the first semester and second semester scores and this is generally reasonable given the shorter length of mid-semester tests and the associated variability of the scores, which is smaller for the mid-semester scores than the first and second semester scores. Thus, in response to the first research question, the results indicate that self-reported scores are accurate or valid from the correlational perspective.

Discrepancy between self-reported and actual scores

Examining the accuracy of the self-reported data from the discrepancy perspective shows that the mean differences between self-reported and actual scores are statistically significant for all nine pairs of scores. In particular, self-reported scores are significantly larger than actual scores for all nine pairs of scores. These findings are consistent with those reported in Sticca et al.'s (2017) longitudinal study of four school subjects (mathematics, German, English, and French) for students in Grades 9-11 in the Swiss school system.

As in the present study, Sticca et al. (2017) found that self-reported grades were significantly higher than the actual grades. This means that in both the present study and that of Sticca et al. (2017), as well as other studies (e.g., Cole & Gonyea, 2010), a larger number of participants, on average, have over-reported their scores. In this study, for example, the proportion of students who over-reported their scores ranged from 43.22% (English mid-semester scores) to a high of 80.34% (for chemistry first semester scores). Overall, the findings show that a very large number of students over-reported their scores.

We can therefore see that even though the correlation between self-reported scores and scores from official school records ranged from moderate to high and was generally acceptable, their discrepancy from official school records is worrisome. Thus, regarding the second research question, the findings indicate that there is a statistically significant difference between self-reported and actual scores for all nine pairs of scores, the former being consistently larger than

the latter. This suggests that findings from studies that exclusively employed self-reported data should be seriously scrutinized, especially when they are conducted to inform decisions about students. For example, if studies are conducted to select low-achieving students (or alternatively high-achieving students) to determine eligibility for a certain program based on the students' self-reported academic achievement scores, such studies are likely to select a number of students who may not meet the intended eligibility criterion. In such situations, it is advisable to use self-reported scores with caution. A better alternative would be to use self-reported scores along with data collected by means other than self-reports (e.g., to use scores from official school records).

Patterns of over-reporting, accurate reporting and underreporting of scores

The findings show a mix of over-reporting, accurate reporting, and underreporting among the students. Despite this general trend, a large number of students over-reported their scores as compared to those students who underreported or accurately reported their scores. While these findings corroborate past research (e.g., Cole & Gonyea, 2010; Sticca et al., 2017), it is noteworthy that the number of students who over-reported their scores in the present study was larger than those who underreported and accurately reported combined on seven out of the nine sets of scores. That is, the number of students who over-reported their achievement scores in the present study was far larger than those reported in previous studies conducted in other contexts (e.g., Sticca et al., 2017).

In addition to this pattern, one can observe that among those who accurately reported their scores, the largest proportion was observed on the mid-semester tests. For each school subject (i.e., English, math, and chemistry), the number of students who accurately reported their scores on the mid-semester tests was larger than the number of students who accurately reported their first- and second-semester scores combined. It should be noted that the time elapsed between the mid-semester tests and data collection for this study was only two to four weeks, and it appears that more students were able to remember and report their mid-semester scores accurately than for either first- or second-semester scores.

Thus, this finding (i.e., more students accurately reported their mid-semester scores than their first- or second-semester scores) may be associated, at least in part, with the relatively short time elapsed between the mid-semester test and the data collection for the present study, which may have facilitated recall of the mid-semester scores. At the same time, it should be noted that

many students have misreported (i.e., over-reported or underreported) their mid-semester scores, which means that the length of time elapsed can partly but not fully account for the students' over-reporting and underreporting behaviors. In other words, other factors could explain the students' over-reporting and underreporting behaviors, but this was beyond the scope of the present study. Future research can therefore focus on identifying other factors that explain why students over-report or underreport their scores.

One can see that the design of this study has provided a possible explanation for why students over-report or underreport their achievement scores when asked to report in surveys. Accordingly, whether the time elapsed is short or long may, at least in part, influence how accurate the students' self-reported scores could be. This can be considered as a contribution of the present study. However, recency cannot account for the entire over-reporting and underreporting behaviors of the participants, which implies the presence of other variables that could be investigated in future research.

In response to the third research question, therefore, the proportion of students who over-reported their achievement scores is significantly larger than the proportion of students who underreported their scores and this is true for all the scores. Besides, among those who accurately reported their scores, the number of students who accurately reported their mid-semester scores is larger than the number of students who accurately reported their first and second semester scores combined.

Differences in accuracy of reporting due to school type and level of academic achievement

The present study found that students from the private school tended to report their scores more accurately than those from the government school did, and this was the case for seven out of the nine scores. To explain this difference, we explored differences between the two groups of students. A closer look at the data and further analysis to examine possible differences in achievement between the students from the two schools yielded significant mean differences in six out of the nine actual scores, and the findings favor students from the private school. More specifically, with the exception of the mid-semester scores, the students from the private school had significantly better actual scores than the students from the government school. That is, the mean actual scores for the first and second semesters of the students from the private school were

significantly better than the corresponding actual scores for the students from the government school.

The study also examined whether students' achievement level is related to the accuracy of self-reported scores. The results indicate that, for all school subjects, the accuracy of reporting is significantly better for students with high achievement scores than for those with low achievement scores. Alternatively, accuracy of self-reported scores is worse among those with low achievement scores than those with high achievement scores. This result corroborates findings from previous research (e.g., Cassady, 2001; Cole & Gonyea, 2010; Mayer et al., 2007). Thus, in the present study as well as in past research, students are more likely to misreport (over-report or underreport) their academic scores, such misreporting is more frequent among those with low achievement scores than those with high and moderate achievement scores.

We should note that the above two findings (pertaining to differences in accuracy of reporting due to school type and achievement level) are in a way related to each other. This is because, as found in the present study, the students from the private school are generally better achievers than those from the government school. Thus, the significant differences in accuracy of reporting due to school type are attributable, at least in part, to the significant difference between the two groups of students in their actual scores collected from official school records.

Conclusion

The study found moderate-to-strong positive associations between self-reported scores and the corresponding scores from school records. In contrast, self-reported scores do not appear to be accurate estimates of actual scores because the mean discrepancies between the paired scores were statistically significant. A disproportionately large number of students (ranging from a low of 43.2% to a high of 80.3%) over-reported their achievement scores. Accuracy of reporting is significantly better among high achievers than among low achievers.

Implications of the findings

The use of self-reported data in research is common. Along with this common use, researchers need to be cautious, particularly if the accuracy of the data is pertinent to their research. For example, if researchers want to study some characteristics of high-achieving students, they should not identify this group of students based merely on self-reported

achievement scores. This is because, as previous studies have consistently shown, many students are likely to over-report their achievement scores, which may lead researchers to consider at least some of them as high achievers when actually they are low or average achievers, making the selected group a mix of low, average and high achievers rather than high achievers only. Thus, in this kind of study, it is important that researchers exercise caution when they use self-reported scores only, and whenever possible it is important that they verify the accuracy of self-reported scores with scores from official school records.

Considering only students who accurately reported their scores, the present study found that many more students accurately reported their mid-semester scores than their first- or second-semester scores. In other words, more students accurately reported their most recent scores (i.e., mid-semester scores) as compared to the ones which they received about a year ago (first semester scores) or about six months ago (second semester scores) at the time of data collection. Based on this, we may consider recency of the scores as a plausible explanation of accuracy of reporting. However, recency of scores does not explain all variation in reporting accuracy, suggesting additional variables may also contribute. Future research can investigate the nature of these variables.

Limitations

One limitation of this study is the small sample size. Although the study findings are consistent with those of previous research conducted elsewhere, future research may examine accuracy of self-reported achievement scores with a larger sample in the Ethiopian context. Besides, the purposive selection of only two (one private and one government) schools may limit generalizability of the findings to other high schools. Furthermore, because the sample of the present study comprises Grade 12 natural science stream students only, the findings may not be generalized to students in lower grades and Grade 12 social science stream students. Future research can therefore investigate the accuracy of self-reported achievement scores among students in lower grades and social science students in Grade 12.

Declaration of Conflicting Interests

No potential conflict of interest was reported by the authors.

Funding information

The authors received no financial support for this research

References

Alebie, H. K., Meshesha, A., Asfaw, A., & Worku, F. (2026). Discrepancies between self-reported measures of hearing with an audiometric measurement of hearing and associated factors in Ethiopia national ear and hearing survey. *Annals of Otology, Rhinology & Laryngology, 135*(2), 115–124. https://doi.org/10.1177/00034894251372825

Cassady, J. C. (2001). Self-reported GPA and SAT: A methodological note. *Practical Assessment, Research, and Evaluation, 7*(1), *Article* 12. https://doi.org/10.7275/5hym-y754

Cole, J. S., & Gonyea, R. M. (2010). Accuracy of self-reported SAT and ACT test scores: Implications for research. *Research in Higher Education, 51*, 305–319. https://doi.org/10.1007/s11162-009-9160-9

Crockett, L. J., Schulenberg, J. E., & Petersen, A. C. (1987). Congruence between objective and self-report data in a sample of young adolescents. *Journal of Adolescent Research, 2*(4), 383–392. https://doi.org/10.1177/074355488724006

Dobbins, G. H., Farh, J.-l., & 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*(4), 321–334. https://doi.org/10.1111/j.1559-1816.1993.tb01090.x

Escribano, C., & Díaz-Morales, J. F. (2014). Are self-reported grades a good estimate of academic achievement? *Studies in Psychology, 35*(1), 168–182. https://doi.org/10.1080/02109395.2014.893650

Fuller, R., Lombardini, S., & Poggi, C. (2023). How accurate are retrospective survey data? Evidence from rural Ethiopia. *Journal of Development Effectiveness 16*(1), 142-158. http://dx.doi.org/10.1080/19439342.2023.2190602.

Gilger, J. W. (1992). Using self-report and parental-report survey data to assess past and present academic achievement of adults and children. *Journal of Applied Developmental Psychology, 13*(2), 235–256. https://doi.org/10.1016/0193-3973(92)90031-C

Hatos, A., & Gyarmati, B. F. (2023). The reliability of self-reported GPA in educational research: A comparison of self-reported and officially recorded data. *Revista Romaneasca pentru Educatie Multidimensionala, 15*(4), 159–177. https://doi.org/10.18662/rrem/15.4/786

Kuncel, N., Crede, M., & Thomas, 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*(1), 63–82. https://doi.org/10.3102/00346543075001063

Maxwell, N. L., & Lopus, J. S. (1994). The Lake Wobegon effect in student self-reported data. *American Economic Review, 84*(2), 201–205.

Mayer, R. E., Stull, A. T., Campbell, J., Almeroth, K., Bimber, B., Chun, D., & Knight, A. (2007). Overestimation bias in self-reported SAT scores. *Educational Psychology Review, 19*(4), 443–454. https://doi.org/10.1007/s10648-006-9034-z

Mehrens, W. A., & Lehmann, I. J. (1991). Measurement and evaluation in education and psychology (4th ed.). Holt, Rinehart & Winston.

Nkhoma, H. H. (2025). The reliability of self-report in clinical psychology: Evidence from Outpatient Clinics in Malawi. *International Journal of Research and Innovation in Social Science, 9*(10), 1125-1132. https://doi.org/10.47772/IJRISS.2025.910000094

Rosen, J. A., Porter, S. R., & Rogers, J. (2017). Understanding student self-reports of academic performance and course-taking behavior. *AERA Open, 3*(2), 1–14. https://doi.org/10.1177/2332858417711427

Shaw, E. J., & Mattern, K. D. (2009). Examining the accuracy of self-reported high school Grade Point Average. College Board Research Report No. 2009-5. Retrieved from https://files.eric.ed.gov/fulltext/ED562616.pdf

Sticca, F., Goetz, T., Bieg, M., Hall, N. C., Eberle, F., & Haag, L. (2017). Examining the accuracy of students' self-reported academic grades from a correlational and a discrepancy

perspective: Evidence from a longitudinal study. *PLoS ONE,* 12(11), e0187367. https://doi.org/10.1371/journal.pone.0187367.

Turner, C. K. (1988). Don't blame memory for people's faulty reports on what influences their judgments. *Personality and Social Psychology Bulletin, 14*(3), 622–629.

Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology*, 95(5), 1202–1216.

Wilson, M. L., & Zietz, J. (2004). Systematic bias in student self–reported data. *Journal for Economic Educators, 4*(4), 13-19.

Zimmerman, M. A., Caldwell, C. H., & Bernat, D. H. (2002). Discrepancy between self-reported and school-reported grade point average: Correlates with psychosocial outcomes among African American adolescents. *Journal of Applied Social Psychology*, *32*(1), 86–109. https://doi.org/10.1111/j.1559-1816.2002.tb01421.x