ALTERNATIVE ESTIMATES OF THE RELIABILITY OF COLLEGE GRADE POINT AVERAGES

Joe L. Saupe

Professor Emeritus

University of Missouri

Mardy T. Eimers

Director, Institutional Research

University of Missouri

Annual Forum of the Association for Institutional Research

June 2 - June 6, 2012

New Orleans, Louisiana

Alternative Estimates of the Reliability of College Grade Point Averages

Joe L. Saupe and Mardy T. Eimers[1]

College grade point averages (GPAs) are used as criteria in the prediction of success in college, as predictors of success in graduate or professional education, as criteria for admission to degree programs, as indicators of qualification for employment, and as variables in different types of research (Warren, 1971). For each of these uses it is important that the GPAs possess some minimum degrees of reliability. For this reason, there have been a number of investigations into the reliability of college grades and GPAs (see Barritt, 1966; Clark, 1950; Etaugh, Etaugh, & Hurd, 1972; Ramist, Lewis, & McCamley, 1990). The reliability of the college GPA has also been used as one variable in studies of some other variable (Bacon & Bean, 2006; Millman, Slovacek, Kulik, & Mitchell, 1983; Singleton & Smith, 1978). An early study (Starch & Elliot, 1913) that dealt with grading high school examinations in mathematics and English indicates there has been interest in the reliability of grades for almost 100 years. Thus, continued contributions to knowledge about the reliability of grades and GPAs and about methods of estimating these types of reliability should be useful. The present study makes an addition to this body of knowledge.

Reliability and College GPAs

Several basic approaches to the estimation of reliability are described in classical measurement theory (Crocker & Algina, 1986; Feldt & Brennan, 1989). The earliest definition

[1] The authors acknowledge and appreciate the assistance of Ann Patton, Senior Programmer, and Nino Kalatozi, PhD candidate in Educational Leadership and Graduate Research Assistant, in the preparation of this paper. They both work in the Office of Institutional Research at the University of Missouri.

of reliability is the correlation between two parallel forms of the same test (Feldt & Brennan, 1989). Test forms are considered to be parallel when they are constructed to cover the same domain or domains of content. It is not clear if there is a counterpart to parallel forms of tests in the case of college GPAs.

A second approach to estimating reliability is the test-retest procedure. With this approach the test is given twice to the same group of subjects and the reliability of the test is estimated by the correlation between the two sets of scores. If two semesters or two years of college coursework are considered to be measures of the same variable, for example academic achievement, then the correlation between GPAs for the two semesters or years may be viewed as a reliability estimate based upon the test-retest situation. Clark (1950) compared correlations between first and second term GPAs with an alternative estimate of the reliability of the GPA for a term. In a second study Clark (1964) examined both approaches to estimating the reliability of GPAs in conjunction with comparing the reliability of grades on an eight-step grading scale with those on a five-step scale. Elliott and Strenta (1988) used correlations among annual GPAs in a study of differences in departmental grading standards. Humphreys (1968) calculated correlations among eight semesters of GPAs. Rogers (1937) also correlated term GPAs for eight academic terms. Werts, Linn, and Jöreskög (1978) used an eight by eight matrix of correlations among semester GPAs in their simplex analysis of that matrix. Finally, Willingham (1985) calculated correlations among yearly GPAs, but did not refer them as reliabilities.

The third type of reliability is estimated by internal consistency methods (Crocker & Algina, 1986). The internal consistency of a test is the degree to which all of the items in the test are measures of the same characteristic or attribute or combination of characteristics or attributes. This type of reliability is estimated on the basis of a single administration of the test. There are at least three different methods that can be used to estimate internal consistency: 1) the split-half procedure, 2) coefficient alpha, and 3) analysis of variance (ANOVA).

The split-half procedure randomly divides the items of a test into two parts and then the correlation between the scores on the two parts is calculated. This correlation is an estimate of the reliability of each half of the test. The estimate of the reliability of the whole test is estimated by use of the Spearman-Brown prophecy formula (Brown, 1910; Spearman, 1910), which expresses the reliability of the total test as a function of the correlation between the two halves of the test. Barritt (1966) used the split- half procedure to estimate the reliability of first semester grades by randomly dividing the grades of students taking 12 or more credits into two sets of courses and correlating the resulting pairs of GPAs. In a similar study involving 38 colleges, Ramist, Lewis and McCamley (1990) randomly divided freshman grades into two halves, calculated correlations between the GPAs of the two halves, and applied the Spearman-Brown formula. The generalized Spearman-Brown formula can be used to estimate the reliability of a test that is three, four, or some greater number times the length of the test for which there is a reliability estimate (Feldt & Brennan, 1989).

A second procedure for estimating the internal consistency type of reliability is known as the coefficient alpha procedure (Cronbach, 1951).[2] The formula for coefficient alpha involves the sum of the variances of the individual item scores and the variance of the total scores on the test. No studies of the reliability of GPAs using Cronbach's alpha were found in the literature reviewed.

Analysis of variance (ANOVA) is a third approach to estimating internal consistency. The most straightforward application of this approach involves a subjects by items analysis of

[2] The Kuder-Richardson formula 20 (Kuder and Richardson, 1937), prominent in the literature on reliability, is equivalent to coefficient alpha when all test items are scored as 0 or 1. This situation does not occur when the measure is a college grade or GPA.

variance (Hoyt, 1941).[3] The reliability estimate is a function of the mean square for students and the interaction or error mean square. Bendig (1953) estimated the reliability of grades for a single course using the ANOVA approach. The course was taught in several sections and four common and one unique instructor-made tests were used.

Other (ANOVA) procedures similar to that of Hoyt are also used. One such procedure is used when some characteristic of a group of subjects is rated, but different raters for different subjects are involved (e.g., Ebel, 1951; Shrout & Fleiss, 1979; Stanley, 1971). For example, Bacon and Bean (2006) used this interclass correlation procedures in their study of the reliabilities of GPAs that differed by number of years included and of GPA in the major versus overall GPA. Etaugh, Etaugh, and Hurd (1972) used the interclass correlation procedure to compare the reliabilities of unweighted mean grades with the reliability of mean grades weighted by their credit values for freshman year and for senior year GPAs. Millman, Slovacek, Kulik, and Mitchell (1983) used the interclass correlation ANOVA procedure to calculate reliabilities of major field GPAs in their study of the effect of grade inflation on the reliability of GPAs.

Other internal consistency procedures for estimating the reliability of GPAs have been suggested. In two previously cited studies Clark (1950) and Clark (1964) investigated the use of a ratio of two standard deviations as the estimate of the reliability of a GPA. Singleton and Smith (1978) calculated the average correlation among the first 20 courses taken by students

---

[3] The reliabilities produced by the coefficient alpha and Hoyt ANOVA formulas are identical and the split-half procedure may be considered to be a special case of the coefficient alpha (Crocker & Algina, 1986). Specifically, the mean of the reliabilities calculated for all possible split halves of a test is very similar to coefficient alpha. The mean is identical to coefficient alpha if the split half is calculated by an alternate formula (Rulon, 1939) that involves differences between the scores on the two half tests rather than the correlation between the half test scores.

and reported the results as reliabilities of individual course grades. The procedures for estimating the reliability of GPAs cited above as illustrations of the test-retest model might also be considered to be members of the internal consistency family.

Researchers who have studied the reliability of GPAs have uniformly used internal consistency procedures. In these studies, because GPA is considered an indicator of overall academic achievement, the internal consistency method is appropriate and it will be employed in the present study.

The literature on the reliability of college grades includes studies of the reliability of individual course grades (Bendig, 1953; Etaugh, Etaugh, & Hurd, 1972; Singleton & Smith, 1978), of single-term GPAs (Barritt, 1966; Clark, 1950; Clark, 1964; Rogers,1937; Werts, Linn, & Jöreskög, 1978;), of one-year GPAs (Bacon & Bean, 2006; Elliott & Strenta, 1988; Etaugh, Etaugh & Hurd, 1972; Humphreys, 1968; Millman, Slovacek, Kulick, & Mitchell, 1983; Ramist, Lewis & McCamley, 1990; Willingham,1985), and of GPAs for more than one year of course work (Bacon & Bean, 2006). There have been relatively few studies of the reliability of the final undergraduate (cumulative) GPA and that GPA is a focus of the present study.

Purposes

The purposes of this study are to focus the attention of researchers and practitioners on the concept of the reliability of the college GPA~~s~~, to provide methods for estimating this reliability, including the method of this study and methods found in the literature, and to add to the body of literature on the topic by examining the following research questions.

1. What are the reliability estimates for one-semester, one-year, two-year and four-year GPAs and how do they differ?
2. How do the results of using the Spearman-Brown formula to estimate the reliabilities of college GPAs compare with coefficient alpha estimates?

3. What is the effect on the reliabilities calculated for multi-semester GPAs of weighting semester GPAs by the credits of those GPAs?
4. How do the reliabilities found in this study compare with similar reliabilities reported in the literature?

In terms of the first research question, previous research suggests that two factors may affect the reliability of GPA over time. In a study of the effects of grade inflation on GPA reliability (Millman, Slovak, Kulik, & Mitchell, 1983), there were non-significant decreases in GPA reliability over time. However, Bacon and Bean (2006) found that four-year overall GPA had a higher reliability (.94) than other limited time frame GPAs, including most recent year (.84) or most recent 2 years (.91). It might be expected that the variance of four-year GPAs is lower than that of first-year GPAs, because of the loss of lower-achieving students between the end of the first year and the end of the fourth year. That lower variance should lead to a lower reliability. On the other hand, adding items to a test can be expected to increase the reliability of the test according to the generalized Spearman-Brown formula (Feldt & Brennan, 1989). In this study a semester GPA is the counterpart of the test item. Thus more semesters should lead to higher reliabilities. Consequently, the exploration of reliability estimates of GPAs at different stages of college completion is of interest.

To address research question 2, the reliabilities of two- four- and eight-semester GPAs are calculated directly and compared to the reliabilities calculated by the generalized Spearman-Brown formula from a one-semester GPA reliability.

The semester GPAs of different students are based upon the differing numbers of credits involved in these GPAs. It might seem that the reliabilities of multi-term GPAs could be improved by giving more weight to those GPAs based upon larger numbers of credits. However, Etaugh, Etaugh and Hurd (1972) found that unweighted GPAs had higher reliabilities than weighted ones. The need for additional information on this matter is the basis of the third research question.

The fourth research question has to do with the possibility of some uniformity among colleges and universities in the patterns of the reliability of cumulative GPAS at different stages in the college experience. Information on this possibility is provided by the comparison of GPAs form the literature with those found in this study.

Following are issues about the reliability of college GPAs that are found in the literature but are not dealt with in this study.

1. That different courses taken by different students may be expected to lead to lower GPA reliabilities than those that would occur if all students take the same courses. In a preceding section of this paper mention is made of the literature on adjusting GPAs for differences in courses taken by different students (Elliott & Strenta, 1988; Young, 1990; Young, 1993).
2. The reliability of the GPA for the courses of a major might be expected to be higher than the overall GPA. However, Bacon and Bean (2006) found that that the opposite is the case.
3. The fact that some students have the same instructor for two terms and others do not may be expected to affect the comparability, hence reliability, of the resulting grades (Clark, 1964).
4. That some students complete more academic terms than others may affect the comparability, hence reliability, of the GPAs (Clark, 1964).
5. The number of points on the grade scale may affect the reliability of GPAs (Komorita & Graham, 1965; Masters, 1974).

Data and Methodology

The data for this study come from a large research university in the Midwest. Specifically, the data are for degree-seeking, full-time and part-time first-time freshman entering the fall semester of 2007, including those who had enrolled for the preceding summer

session. There were 4,970 students in this entering class. Forty-seven of these students did not remain enrolled long enough for an academic record to be posted for them at the end of that initial semester. End-of–semester credits and semester GPAs are recorded for each student for each of the eight semesters. Summer session and intersession GPAs are not included. The numbers of consecutive semesters the 4,970 students remained enrolled as well as the students' cumulative GPAs at the end of the first, second, and fourth years as recorded in university records are included.

From these data, cumulative GPAs for the end of the first two, first four, and all eight semesters are calculated, as well as weighted semester GPAs for students completing two, four, eight semesters. A weighted GPA was calculated by multiplying the semester GPA by the ratio of the number of credits in that GPA by the mean number of credits in the GPAs of all students for that semester.

The reliabilities calculated from the semester GPAs are the reliabilities of the sums or the means of the GPAs for the included semesters. These mean GPAs are not identical to the true cumulative GPAs that are recorded in the student's academic records. These GPAs involve the semester-by-semester numbers of credits completed. The reliabilities of the sums or means of the weighted semester GPAs may be better estimates of the reliabilities of the true cumulative GPAs and this is the reason weighted semester GPAs are calculated and included in the study.

The following data analyses are carried out:

- Correlations among the following GPAs are calculated for students completing two, four and eight semesters:
    1. Actual cumulative GPAs
    2. Cumulative GPAs calculated from the semester GPA data
    3. Means of semester GPAs
    4. Means of weighted semester GPAs

These correlations are calculated in order to determine the degree to which they are interchangeable. Specifically, do the means of semester GPAs accurately reflect the cumulative GPAs; do the calculated cumulative GPAs that exclude summer and intersession data accurately reflect the actual cumulative GPAs; and how are the means of the weighted GPAs related to the other three measures?

- Correlations between semester-one and semester-two GPAs and between weighted semester-one and semester-two GPAs are calculated in order to estimate the reliability of first-year, one-semester GPAs and to compare this reliability for unweighted and weighted GPAs.
- Internal consistency reliabilities are calculated using Cronbach alpha (Cronbach, 1951) for end of two-semester, end of four-semester and end of eight-semester mean GPAs, unweighted and weighted, in order to compare GPA reliabilities over time and to compare reliabilities of unweighted and weighted GPAs. Using symbols for the GPA, the formula is: alpha = $\frac{s}{s-1}(1-\frac{VARsem}{VARgpa})$, where s is the number of semesters, VARsem is the variance of GPAs for a semester, and VARgpa is the variance of the sums of GPAs.
- Based on the reliability of one-semester GPAs, the Spearman-Brown procedure (Spearman, 1910; Brown, 1910) is used to estimate the reliability of two-semester, GPAs, four-semester GPAs, and of eight-semester GPAs in order to compare the several procedures for estimated the reliability of the several comparable GPAs. The basic Spearman-Brown formula for estimating the reliability of a two-semester GPA is: SB = $\frac{2r}{1+r}$, where r is the correlation between the two semester GPAs. The generalized formula for estimating the reliability of a four- or eight-semester reliability is: GSB = $\frac{sr}{1+(s-1)r}$, where s is the number of semesters for which the reliability is to be estimated.

Results

The data analyses are carried out on groups of students defined on the basis of the number of consecutive semesters they completed. This basis for the selection of students to be included in an analysis is used so that all students included in a calculation of reliability had completed the same number of semesters without gaps in their attendance. Where there are gaps in the sequences of semesters completed, the coefficient alpha procedure would not be applicable. The alpha procedure allows differences among semesters to be ignored in the estimation of the reliability of the sum or mean of semester GPAs.

Table 1 shows the numbers and cumulative numbers of students completing each of the consecutive number of semesters. From the cumulative numbers, 4,606 students are included in the analyses of students completing two consecutive semesters, 3,962 in the analyses for students completing four consecutive semesters, and 2,968 in those for students completing eight consecutive semesters.

Insert Table 1 about here

Table 2 contains the correlations among the four cumulative or mean GPAs for the groups of students completing two, four, and eight consecutive semesters. Means and standard deviations of the four overall GPAs are included for each group. The correlations among the actual cumulative GPAs, the calculated cumulative GPAs, and the mean GPAs exceed .99 for all three groups of students. The means and standard deviations for these three overall GPAs are comparable within each of the three groups with the mean of the actual cumulative GPA slightly, but consistently exceeding the means for the other two measures. Also, the standard deviations for the actual cumulative GPAs are slightly, but consistently smaller than those for the other overall GPAs.

The correlations of the means of weighted GPAs with the other three overall GPAs are consistently smaller than the intercorrelations among the first three overall GPAs. While the

means of these GPAs are comparable to the means of the first three GPAs, their standard deviations are appreciably higher.

Insert Table 2 about here

The mean GPAs increase and the standard deviations decrease as the number of semesters included increases. These trends are not surprising. In addition to possibly differing grading standards between freshman or sophomore courses and courses taken by juniors and seniors, these trends very likely reflect the loss of lower achieving students over four years of the study.

Table 3 provides the several reliability estimates for one-semester, two-semester, four-semester, and eight-semester GPAs. The one-semester reliabilities are correlations between first- and second-semester GPAs for students who completed the first two semesters. The Spearman-Brown estimates are derived from the one-semester reliabilities in the table. The remaining reliabilities are coefficient alphas calculated for each group of students completing two- four- or eight-consecutive semesters. The one-semester reliabilities, .72 and .69, are similar, but the value for unweighted GPAs is modestly higher than the one for weighted GPAs. The Spearman-Brown values for two-, four-, and eight-semester unweighted and weighted GPAs are also similar with differences ranging from .02 to .00. The alpha reliabilities for unweighted GPAs consistently, but modestly, exceed those for weighted GPAs. The Spearman-Brown estimates for four- and eight-semester GPAs are moderately higher than the corresponding alphas. Finally, in each case the reliability estimate increases from approximately .70 to .91 or higher as the number of semesters increase.

Insert Table 3 about here

Reliabilities for one-, two-, four-, and eight- semester GPAs from the literature that are comparable to those of this study, including those found in this study are as follows:

- One-semester: .72 (this study), .84 (Barritt, 1966), .70 (Clark, 1964), .66 (Humprheys, 1968), and .80 (Rogers, 1937).
- Two-semester: .84 (this study), .84 (Bacon & Bean, 2006), .69 (Elliott and Strenta, 1988), .81 (Etaugh, Etaugh, & Hurd, 1972), .83 (Millman, Slovacek, Kulik, & Mitchell, 1983), .82 for (Ramist, Lewis, & McCamley, 1990), and .70 (Willingham, 1985).
- Four-semester: .86 (this study) and .90 (Bacon and Bean, 2006).
- Eight-semester: .91 (this study) and .91 (Bacon and Bean, 2006).

Other reliabilities of GPAs are reported in the literature, but the ones above are the ones most comparable to those of this study. A few decisions had to be made to select these comparable reliabilities. For example, in a couple of cases the average of two or more reliabilities from a single study is used. Also, the one-semester reliabilities used here are first-semester (or second-semester) reliabilities; values for subsequent semesters are not selected.

To facilitate comparisons of these reliabilities Chart 1 is provided. The chart shows the relationship between the number of semesters, one through eight, of coursework on which a GPA is based, and the reliability of that GPA. The values in the chart are given above.

Insert Chart 1 about here

These reliabilities were derived using a variety of procedures. This study is the only one that made use of coefficient alpha. The split-half procedure and the Spearman-Brown formula are used in this study and in others. Other studies employed various analysis-of-variance approaches to estimating GPA reliability. It might be expected that values of reliabilities estimated by different procedures would to some degree be dependent on the procedure used. Also, the studies were carried out with data from a variety of colleges and universities. The reliability of a GPA might be expected to vary from one type of institution to another. For example, the university from which the data of this study comes is comprehensive, offering a

great variety of undergraduate majors. To the degree that grading standards vary to some degree among majors this variety of majors might be expected to depress the reliability of overall GPAs. Thus, Chart 1 should be considered preliminary and not definitive. It does suggest there is a generally positive relationship between the two variables.

Discussion

As previously noted the alpha reliabilities of this study are the reliabilities of sums of semester GPAs. They correspond to the total scores on a test for which an alpha is calculated. To make these sums of GPAs comparable to other GPAs they are divided by the appropriate number of semesters and expressed as means. Also, these means of semester GPAs exclude grades earned in summer sessions or intersessions. The GPAs that should be of interest are the cumulative GPAs that appear in the student's official records. These GPAs are, of course, influenced by the numbers of credits on which each semester GPA is based and include grades earned in summer sessions and intersessions. The correlations, over .99, between the means of semester GPAs and the actual cumulative GPAs and the similarity of the means and standard deviations of these two variables indicate that the alpha reliabilities of this study are very good estimates of the reliabilities of the cumulative GPAs in the student's records. The third indicator of overall achievement, the cumulative GPA calculated from semester GPAs and credits, also excludes grades earned in summer sessions and intersessions and is included in the study in order to discern if the exclusion of these grades impacts the accuracy of the alpha reliabilities. The near 1.00 correlations among these three overall GPAs and the similarity of their means and standard deviations suggest they are essentially interchangeable and provide confidence that the alpha reliabilities are very good estimate of the reliabilities of the actual cumulative GPAs

The lower correlations involving the means of weighted GPAs and the higher standard deviations for these variables indicate that the weighting procedure does not improve the comparability of these overall GPAs to the actual cumulative GPAs. As a matter of fact, the weighting procedure distorts the validity of the resulting GPAs. This finding is reinforced by the fact that the reliabilities of the GPAs resulting from the weighting procedure are lower than the reliabilities of the corresponding unweighted GPAs. Etaugh, Etaugh, and Hurd (1972) also found that weighting GPAs results in lower reliabilities for composite GPAs than does not weighting.

The one semester reliabilities of .72 (unweighted) and .69 (weighted) are correlations between semester one and semester two GPAs. The Spearman-Brown values for two-semester GPAs are the results of applying the basic Spearman-Brown formula to the respective correlations and the Spearman-Brown values for two-, four-, and eight-semester GPAs are products of the generalized Spearman-Brown formula. The similarity of the two reliabilities for two-semester GPAs and of the six reliabilities for two-, four-, and eight-semester GPAs indicates that the Spearman-Brown technique, as applied here, produces quite reasonable estimates of the reliabilities of GPAs for more than one semester of coursework. That the reliabilities of weighted GPAs are consistently lower than the reliabilities of unweighted GPAs is another indication that the weighting procedure is undesirable. The conclusion must be that the weighting procedure contributes error variance to the resulting average GPAs. In other words, it decreases the validity of the overall GPAs as indicators of student's academic achievement.

The Spearman-Brown estimates of reliabilities for four- and eight-semester GPAs exceed their corresponding alpha reliabilities. Although the differences are not large, this result suggests that the alpha reliabilities are affected by the decrease in the variances of overall GPAs as the number of semesters increase. The Spearman-Brown estimates are not affected by this decrease in variance.

Reliabilities of GPAs found in this study are not unlike those taken from then literature. For the five one-semester GPAs the range is from .66 to .84 (.72 from this study). Seven two-

semester GPA reliabilities range from .69 to .84 (.84 in this study). There are only two four-semester reliabilities, .90, and from this study, .86, and two- eight–semester values, .90 and from this study, .91. There are clearly too few values for a Meta analysis of these values, but a trend in the relationship between the reliability of the GPA and the number of semesters on which it is based is suggested by these data. As portrayed by the line fitted in Chart 1 the GPA reliability increases at a decreasing rate as the number of semesters increases. Additional research is needed to confirm this relationship.

The reliability of a GPA determines an upper bound to the correlation of that GPA with another variable. If the GPA were perfectly reliable, the correlation would be higher than that observed with the GPA that has a reliability of less than 1.00. For example, Saupe and Eimers (2011) in a study of how restriction of range in high school GPA depresses correlations in the prediction of success in college note that unreliability in the college success variable is another factor that depresses such correlations. They found a correlation of .56 between high school core course GPA (CCGPA) and freshman year GPA (FYGPA). If the reliability of the FYGPA is .84, as found in the present study, then using the relationship[4] provided by Walker and Lev (1953), the correlation between CCGPA and a perfectly reliable FYGPA would be .61.

Conclusions

The following conclusions seem warranted:

[4] $R^c_{hc} = R_{hc} / \sqrt{R_{cc}}$, where R_{hc} is the original correlation between HSGPA and FYGPA, R_{cc} is the reliability of FYGPA and R^c_{hc} is the estimated correlation between HSGPA and FYGPA assuming the reliability of FYGPA is 1.00.

1. Means of semester GPAs are almost identical to actual cumulative GPAs. Consequently, the reliabilities of sums (or means) of semester GPAs are good estimates of the reliabilities of actual cumulative GPAs.
2. Reliabilities of cumulative GPAs increase from the first semester to the end of the undergraduate program at a decreasing rate. In the present study, the increase is from .72 for the first-semester GPA, .84 for the two-semester GPA, and .86 for the four-semester GPA ,to .91 for the eight-semester or near-final undergraduate GPA. Similar values and trends are likely to be found at other colleges and universities.
3. The use of the Spearman-Brown generalized formula to estimate reliabilities of longer term GPAs from the reliability of first-semester GPA provide generally accurate, but moderately overstated, values.
4. Reliabilities calculated from weighted semester GPAs understate the reliabilities calculated from unweighted GPAs and weighted GPAs do not provide good estimates of actual cumulative GPAs.

Implications

The true standard of academic success is represented by a student's grade point average. Whether the GPA is cumulative, by semester, or calculated in some other manner—it is critically important. The GPA can impact a college student's ability to pursue that coveted major, maintain or qualify for a financial aid award or scholarship, get into the graduate school of choice, or land the job that propels the graduate to greater opportunities. As easily as it can open doors, GPA thresholds can also keep students out. Consequently, it is important to know as much about the GPA as possible: including its reliability.

The purpose of this study is to examine the reliability of college GPAs, to provide different methods for estimating these reliabilities, and to add to the knowledge base in terms of the research literature and practical application in colleges and universities. Thus, the following implications are proposed. First, the user of college GPAs should be aware of the fact

that the reliabilities GPAs vary according to the stage of the college career at which the GPAs are determined. It appears that the reliability increases as additional coursework in completed Also, it can be expected that even as early as the end of the first year the reliability of the GPA may well be at an acceptable level of .80 or higher.

Second, there are a number of methods that can be used to estimate the reliability of a college GPA. This study introduced coefficient alpha as a method for determining the reliability of a GPA. This method may prove to be beneficial to institutional researchers and faculty researchers who examine the reliability of college GPAs.

Third, frequently researchers and practitioners alike do not think about the reliability of college GPA. They may be interested in understanding how well admission tests (e.g., ACT, SAT, etc.), high school rank in class, high school GPA, and similar variables predict success in college. Success in college is almost always tied to the student's GPA in some manner. However, how often is the reliability of the dependent variable, the GPA considered? How often is the reliability of the GPA at different periods of time over a student's career questioned? If this study has highlighted the importance of GPA reliability in both practical and scholarly pursuits, it will have accomplished a principal goal.

References

Bacon, D. R., & Bean, B. (2006). GPA in research studies: an invaluable but overlooked opportunity. *Journal of Marketing Education*, *28*(1), 35-42.

Barritt, L. S. (1966). The consistency of first-semester grade point average. *Journal of Educational Measurement, 3*(3), 261-262.

Bendig, A. W. (1953). The reliability of letter grades. *Educational and Psychological Measurement, 13*(2), 311-321.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, *3*(3), 296-322.

Clark, E. L. (1964). Reliability of grade point averages. *The Journal of Educational Research, 57*(8), 428-430.

Clark, E. L. (1950). Reliability of college grades. *American Psychologist,* 5(7), 344.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York, NY: Wadsworth Publishing.

Cronbach, L. J. (1951). Coefficient alpha and the internal consistency of tests. *Psychometrika, 16*(3), 297-334.

Ebel, R. L. (1951). Estimation of the reliability of ratings. *Psychometrica*, 16(4), 407-424.

Elliott, R., & Strenta, A.C. (1988). Effects of improving the reliability of the GPA on prediction generally and on comparative predictions for gender and race. *Journal of Educational Measurement, 25*(4), 333-347.

Etaugh, A. F., Etaugh, C.F., & Hurd, D. E. (1972). Reliability of college grades and grade point averages: some implications for the prediction of academic performance. *Educational and Psychological Measurement, 32*(4), 1045-1050.

Feldt, L. S., & Brennen, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (pp. 105-146). Washington, DC: American Council on Education.

Hoyt, C. (1941). Test reliability estimated by analysis of variance. *Psychometrika, 6*, 153-160.

Humphreys, L. G. (1968). The fleeting nature of the prediction of college academic success. *Journal of Educational Psychology*, 59( ), 375-380.

Komorita, S. S., & Graham, W. K. (1965). Number of scale points and the reliability of the scales. *Educational and Psychological Measurement, 4,* 987-995.

Kuder. G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2*, 151-160.

Masters, J. R. (1974). The relationship between number of response categories and reliability of Likert-type questionnaires. *Journal of Educational Measurement, 11,* 49-53.

Millman, J., Slovacek, S. P., Kulick, E., & Mitchell, K. J. (1983). Does grade inflation affect the reliability of grades? *Research in Higher Education, 19*(4), 423-429.

Ramist, L., Lewis, C., & McCamley, L. (1990). Implications of using freshman GPA as the

criterion for the predictive validity of the SAT. In W. W. Willingham, C. Lewis, R. Morgan, & L. Ramist (Eds.), *Predicting college grades: An analysis of institutional trends over two decades* (pp. 253-288). Princeton, NJ: Educational Testing Service.

Rogers, H. H. (1937). The reliability of college grades. *School and Society, 45,* 758-760.

Rulon, P. J. (1939). A simplified procedure for determining the reliability of a test by split-halves. *Harvard Educational Review, 9*(1), 99-103.

Saupe, J. L. & Eimers, M. T. (2011). Correcting correlations when predicting success in college. *IR Applications,* 31, 1-11.

Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlation: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420-428.

Singleton, R., & Smith, E. R. (1978). Does grade inflation decrease the reliability of grades? *Journal of Educational Measurement*, *15*(1), 37-41.

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology,* 3(3), 271- 295.

Stanley, J. C. (1971). Reliability. In R. L Thorndike (Ed.), *Educational measurement* (pp. 356-442). Washington, DC: American Council on Education.

Starch, D., & Elliot, E. C. (1913). Reliability of grading work in mathematics. *School Review, 21*, 294-295.

Walker, H. M., & Lev, J. (1953). *Statistical Inference*. New York: Henry Holt.

Warren, J. R. (1971). *College grading practices: An overview*. Washington, D. C.: ERIC Clearinghouse on Higher Education.

Werts, C., Linn, R.L., & Jöreskög, K.G. (1978). Reliability of college grades from longitudinal data. *Educational and Psychological Measurement, 38*(1), 89-95.

Willingham, W. W. (1985). *Success in college*. New York, NY: College Entrance Examination Board.

Young, J. E. (1990). Are validity coefficients understated due to correctable defects in the GPA? *Research in Higher Education, 31(4),* 319-325.

Young, J. E. (1993). Grade adjustment methods. *Review of Educational Research, 63*(2), 151-165.

Table 1

Numbers and Percentages of Students Who Completed Given Numbers of Consecutive Semesters

Consecutive Semesters	Number of Students	Cumulative Number	Percent of Students	Cumulative Percent
8	2,968	2,968	59.7%	59.7%
7	290	3,258	5.8%	65.6%
6	228	3,486	4.6%	70.1%
5	183	3,669	3.7%	73.8%
4	253	3,922	5.1%	78.9%
3	206	4,128	4.1%	83.1%
2	478	4,606	9.6%	92.7%
1	317	4,923	6.4%	99.1%
0	47	4,970	0.9%	100.0%
Total	4,970	--	100.0%	--

Table 2

Correlations Among and Means and Standard Deviations of the Four Cumulative or Mean Two, Four, and Eight Semester GPAs

Variable	Calculated Cum GPA[2]	Mean of Sem GPAs[3]	Mean of Whtd Sem GPAs[4]	Mean	S.D.
Two-Semester GPAs (N = 4,606)					
Actual Cum GPA[1]	0.996	0.994	0.941	2.95	0.74
Calculated Cum GPA[2]		0.998	0.945	2.94	0.75
Mean of Sem GPAs[3]			0.944	2.94	0.75
Mean of Whtd Sem GPAs[4]			---	2.97	0.88
Four-Semester GPAs (N = 3,922)					
Actual Cum GPA[1]	0.994	0.993	0.934	3.10	0.55
Calculated Cum GPA[2]		0.998	0.938	3.08	0.57
Mean of Sem GPAs[3]			0.937	3.08	0.57
Mean of Whtd Sem GPAs[4]			---	3.10	0.70
Eight-Semester GPAs (N = 2,968)					
Actual Cum GPA[1]	0.994	0.992	0.916	3.19	0.46
Calculated Cum GPA[2]		0.998	0.921	3.16	0.49
Mean of Sem GPAs[3]			0.920	3.16	0.49
Mean of Whtd Sem GPAs[4]			---	3.17	0.58

[1] Cumulative GPA take from the University's student data base.

[2] Calculated cumulative GPA from semester GPAs and credits.

[3] Mean of semester GPAs.

[4] Mean of weighted semester GPAs.

Table 3

Internal Consistency Reliability Estimates by Number of Semesters and Method of Estimating Reliability

Method of Reliability Estimate	One Semester	Two Semesters	Four Semesters	Eight Semesters
N	4,606	4,606	3,922	2,968
Correlation - Unweighted GPAs	0.72	--	--	--
- Spearman-Brown	--	0.84	0.91	0.95
Correlation - Weighted GPAs	0.69			
- Spearman-Brown		0.82	0.90	0.95
Alpha - Unweighted GPAs	--	0.84	0.86	0.91
Alpha - Weighted GPAs	--	0.81	0.85	0.85

Chart 1

Reliability of GPA by Number of Semesters, Data from This Study and from the Literature