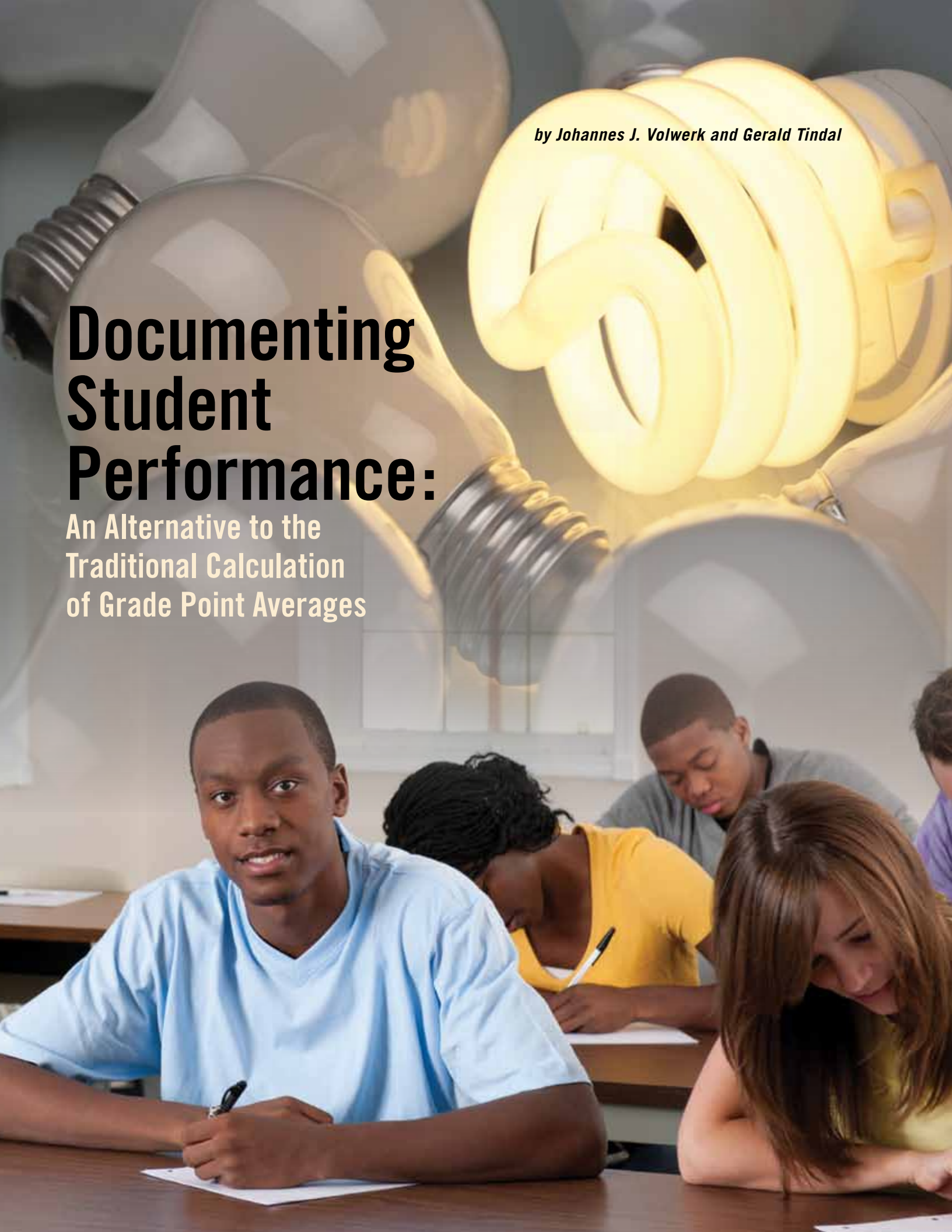


by Johannes J. Volwerk and Gerald Tindal

Documenting Student Performance:

An Alternative to the Traditional Calculation of Grade Point Averages

Abstract

Traditionally, students in secondary and postsecondary education have grade point averages (GPA) calculated, and a cumulative GPA computed to summarize overall performance at their institutions. GPAs are used for acknowledgement and awards, as partial evidence for admission to other institutions (colleges and universities), and for awarding scholarships. Given the high stakes nature of the GPA as a tool to compare and rank overall student proficiency and potential, it is important that it be appropriately scaled. In this study, results from a re-scaled GPA reflect a true (proportional) interval scale and are compared to the traditional (ordinal) GPAs for various student groups thereby demonstrating that differential effects for critical populations are absent so that equality opportunity is maintained. Implications, including, lack of research on this topic are discussed.

Introduction

The grade point average (GPA) plays a significant role in the assessment of a student's overall past academic achievements and future potential for such purposes as college admission, admission to graduate programs, awarding of scholarships, and entry into training programs and the workforce. Although a variety of other measures and outcomes also may come into play for these purposes, GPA is often looked at first because it is assumed to express a student's ability and future potential in a simple, numerical and easily comparable way.

Many universities have established GPA admission criteria. For example, at the University of Oregon, the standard admission requirements are listed as (a) high school GPA of at least 3.00, (b) graduate from a standard or regionally accredited high school, (c) a grade of C- or higher in 14 college preparatory courses, and (d) SAT or ACT scores (http://admissions.uoregon.edu/freshmen/requirements). At some universities, automatic admission may simply be based on a student's GPA meeting a minimum threshold value, with other factors being considered only if the GPA is below the minimum value. A newspaper article about last year's PAC-10 conference football coach of the year reported that the amount of a bonus payment to the coach was directly tied to the team GPA for scholarship players: $25,000 for a team GPA above 2.5; $50,000 for a GPA above 2.75; and $100,000 for a GPA above 3.0 (Mosely 2010).

The critical importance (and value) of GPAs is not confined only to admission and awards, however. In the world of research, equal importance is placed on GPAs as a viable outcome measure (dependent variable). For example, Fewster and Macmillan (2002) used GPAs in their study of curriculum-based measurement: "Schools generated student grades for junior secondary school courses in a number of different formats. Course grades were generally provided as year-end percentages, although some were reported as letter grades… English and social studies course grades were chosen as criterion measures because teacher experience indicates that these courses are the most reading and writing intensive" (p. 152).

The implicit assumption in using GPA values for these purposes is that it is a faithful measurement of a student's overall achievement level in secondary or postsecondary education and, by extrapolation, a reliable predictor of future success. However, this study shows a number of problems with GPAs calculated in the traditional manner, and argues that a better measurement of overall classroom performance is possible by abandoning letter grades and grade points and using a different scale of overall classroom assessment.

The crux of the problem with the GPA calculation, as shown by the GPA line, is that averaging of letter grades with the traditional way of assigning grade points results in a disproportional lowering of the GPAs for any student who does not earn an "A" grade in every class.

Traditional GPA Calculation

Calculation of a student's GPA both in secondary and higher education almost universally involves a process such as summarized below.

Classroom Percentage	Letter Grade	Grade Points Awarded
90-100	A	4
80-90	B	3
70-80	C	2
60-70	D	1
0-60	F	0

Classroom percentages are obtained by averaging over a variety of assignments and result in a corresponding letter grade being recorded in the student's transcript. The GPA is then obtained by calculating a weighted average of the grade points associated with each letter grade using the (potential) credit that the student earned for each course, semester, or term as the weighting factor. The cutoff for earning credit towards high school graduation is usually set at 60 percent (D letter grade) but four-year colleges may set a higher standard (70 percent or C letter grade) for a course to count towards college admission. In this system students who earn a failing grade not only receive no credit for the course, they also receive zero grade points towards the calculation of their GPA.

Some variations on this general theme may exist. For example, teachers may use slightly different percentage ranges to assign particular letter grades or may use different rounding methods. More difficult courses, for example, honors or Advanced Placement (AP), may be weighted differently, or additional fractional grade points (for example, 0.3) may be added for scores at the high end of the percentage range (A+, B+, etc.) and subtracted for scores at the low end of the percentage range (A-, B-, etc.). Because these practices may vary from school to school and district to district they compromise the reliability of the GPA as a facile comparative tool for measuring overall student achievement.

The traditional method of GPA calculation, however, presents a number of other problems best discussed using Figure 1 as an illustration. In Figure 1, the discontinuous line graph shows the relationship between the classroom percentage earned and the grade point awarded based on the letter grade assigned by the teacher. The graph was constructed using the percentage range intervals shown above. Adding or subtracting fractional grade points for + or – letter grades would increase the number of "steps" in the right hand part of the graph but this does not affect the basic arguments put forward.

The crux of the problem with the GPA calculation, as shown by the GPA line, is that averaging of letter grades with the traditional way of assigning grade points results in a disproportional lowering of the GPAs for any student who does not earn an "A" grade in every class. The more lower letter grades (B, C, D, F) are present in the student's transcript, the more significant is the disproportional lowering thus increasing the difference between high and low achieving students purely as the result of a calculation artifact.

Other issues arise. First, there is no room at the top. Because any score of 90 percent and above results in 4.0 grade points being awarded, the GPA does not distinguish between a student who consistently scores in the higher 90s and a student who consistently scores in the lower 90s. In other words, the GPA allows no distinction between a truly outstanding student and one who is very good. In many schools, class ranking has been abandoned because of the over-abundance of 4.0 GPAs, even though this information is still of interest to many institutions using GPA as one of their evaluation tools. This same problem exists for the other percentage ranges. Information about a student's overall performance is lost because, in the process of calculating a GPA, a transition is made from a linear (continuous) scale (classroom percentage) to a nonlinear (ordinal) scale (letter grade and grade point).

A second problem is that minor changes in the classroom percentages earned by students may result in exaggerated differences in the grade points awarded and consequently in the calculated GPA. A hypothetical example illustrates this point. Student A takes two courses and earns 90.0 percent of the points in both courses. She is awarded two A letter grades and four grade points for each course with a resulting GPA of 4.0. Student B takes the same two courses and earns 89.9 percent in one and 90.0 percent in the other. She is awarded a B and an A, 3.0 and 4.0 grade points, respectively, with a resulting GPA of 3.5. A fraction of a percent difference in the classroom percentage may be the difference between guessing right or wrong on a single multiple choice question on a test but it results in a significant difference in the GPAs for students who, based on their classroom percentages, essentially achieved at an equal level.

The third and major problem is that there is too much room at the bottom. The grade point awarded for any classroom percentage below 60 percent is zero. Therefore, in the calculation of the GPA, no distinction is made between a student earning 10 percent in a class (turned in a minimal amount of work, failed all tests) and a student earning 59.9 percent (turned in all work but did poorly on some tests). In this system failure is failure and there is no incentive for students who may foresee a failing grade to keep working hard to earn the best percentage they can.

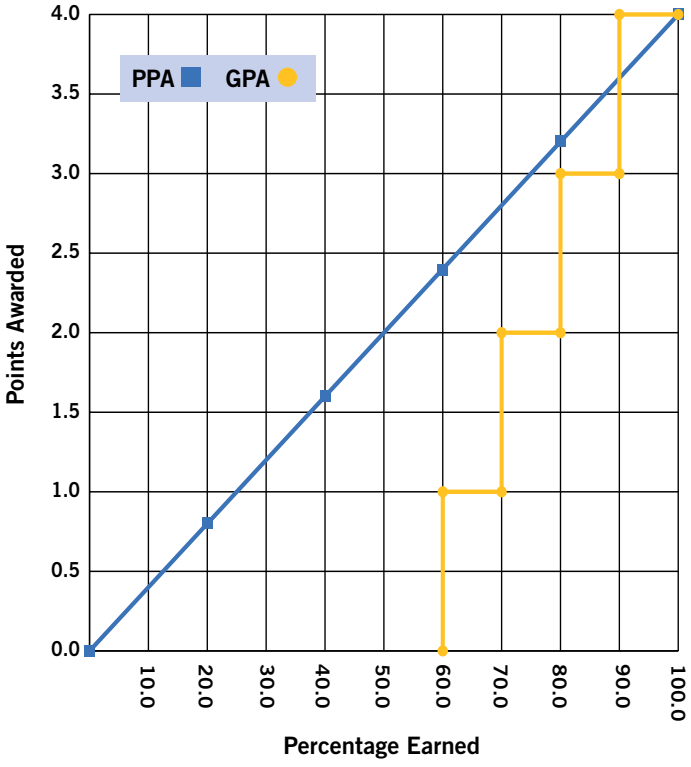
In summary, the traditional method for calculating GPAs tends to inflate GPAs at the top of the four-point scale and deflate GPAs at the bottom of the four-point scale. It also tends to both exaggerate and mask differences in student achievement. The nonlinear relationship between classroom percentages and grade points awarded tends to widen the gap between higher and lower achieving students and thus presents an unfair disadvantage to the latter. These tendencies more severely compromise the use of GPA as a straightforward tool to evaluate and compare students' achievements and abilities, and predict their future potential.

Percentage Point Averages: A Logical Alternative

This study proposes an alternative method to the traditional GPA as a measurement of a student's overall classroom/course achievement that circumvents the problems described above. The proposal is to abolish letter grades and the grade points associated with these letter grades, and utilize the classroom percentages directly to calculate an alternative to the GPA, termed "percentage point average" (PPA). In this procedure, a classroom percentage is simply converted to percentage points awarded using a linear scale as illustrated by the straight line in Figure 1. To facilitate easy comparison between GPA, which usually spans a four-point range, and PPA values, percentages are converted to the corresponding values on a 0-4 point scale. However, in principle any percentage point scale (for example, 0-10, 0-100, 0-7) could be used.

The points awarded form a continuous range and are directly proportional to the earned classroom percentages. There is no disproportional lowering of the measure of student achievement, because averaging of any number of points on the straight line in Figure 1 produces a point that also is on the line. The PPA is calculated as the weighted average of the assigned percentage points with, as for the GPA, using (potential) credit as the weighting factors. Alternatively, a weighted average of the percentages can be calculated first and then converted to a four-point scale (or other scale). These procedures are mathematically equivalent and yield the same result.

Figure 1. Relationship Between Earned Classroom Percentage and Grade Points Awarded



Calculating a student's PPA rather than a GPA has other advantages when juxtaposed to other problems with the GPA calculation as described above. Because of the direct linear relationship

One of the main effects of the switch from GPA to PPA is that it brings up the low end. Lower achieving students, those earning F, D and C letter grades, and whose averages are disproportionately lowered because of the nonlinear relationship between percentage and grade point…

between classroom percentage and percentage point the "bunching," distortion, and loss of information associated with the traditional method does not happen. In the example used above, where students A and B had almost identical classroom percentages, but ended up with GPAs of 4.0 and 3.5, respectively, their PPAs calculate as 3.600 and 3.598. These values are essentially identical in line with their classroom performances. When letter grades are abolished, teachers will no longer have to agonize over "finding" a few tenths of a percent so that a student can get the higher letter grade thereby reducing the chance of "undue influence" or "questionable practices" in classroom assessment.

A second advantage is that now there is room at the top. The percentage range (90–100 percent) that would result in identical GPAs of 4.0 now allows distinction between these high achieving students because the 90–100 percent range translates into a PPA range of 3.60–4.00. As a consequence, class ranking based on overall student achievement can be restored in those schools where it has been abandoned because of the over-abundance of "four points." In fact, when PPA is adopted, a "four-point" student will become exceedingly rare, as it requires a 100 percent average classroom percentage. Organizations now "blindly" awarding scholarships based on 4.0 GPAs may need to refine their policies and make use of the additional information the PPA provides.

The most important advantage resulting from the PPA method is how it affects lower achieving students. No longer will students earning failing grades receive the "double whammy" of not earning credit and getting zero grade points, irrespective of whether they "missed by a mile" or just barely failed the course. In the PPA calculation, every classroom percentage score counts for itself even when that score is below the cutoff limit for earning credit. This is a great incentive for students who find themselves below the 60 percent limit in a class to not give up but continue to do the best they can because in terms of their overall achievement the higher their percentage the higher their PPA. This should also be a great incentive for teachers to encourage their students to do just that, and may reduce the behavioral problems often observed with students who have given up and become disengaged.

One of the main effects of the switch from GPA to PPA is that it brings up the low end. Lower achieving students, those earning F, D and C letter grades, and whose averages are disproportionately lowered because of the nonlinear relationship between percentage and grade point, stand to benefit the most from having a PPA as their measure of overall classroom achievement. Widening of the gap between higher and lower performing students, an artifact of the GPA calculating procedure, no longer occurs. Because student demographics such as racial or ethnic minority, male, special education, and low income often are associated with lower achievement, the so-called "achievement gap," one would expect that those groups specifically would benefit from having a PPA rather than a GPA on their transcripts. In other words, a better measurement of their overall performance would tend to narrow the achievement gap by removing the artificial widening that results from the GPA calculating procedure.

The High School GPA Project

To test the feasibility of the PPA as an alternative to the GPA, a pilot longitudinal study at a high school in a Northwest city was conducted, as one researcher was a teacher in the science department at this school.

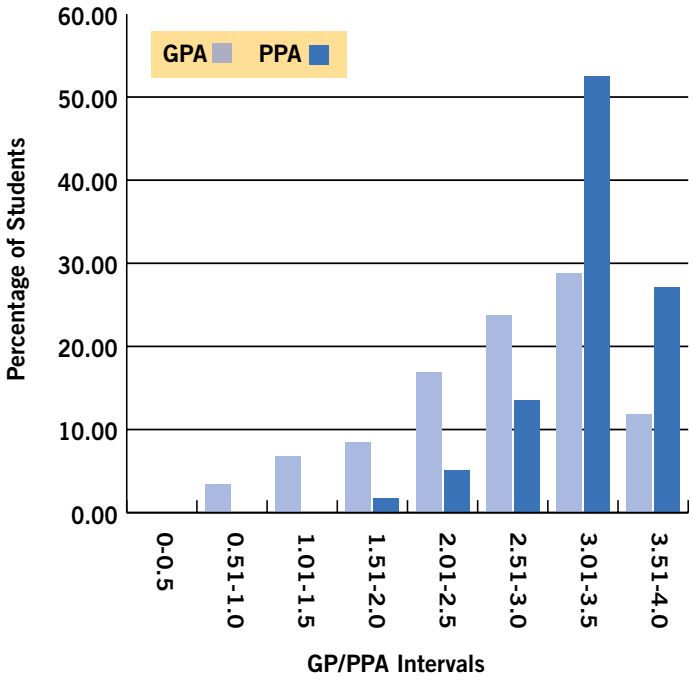
Briefly, the project encompassed the following:
Out of the 2008–2009 class of incoming freshmen into the high school comprehensive program, a cohort of 60 students was randomly selected. For each of these students, teachers in the core areas (science, math, language arts, and social studies) were asked to report letter grades, classroom percentages, and credit earned (potential credit if the student failed the class) at the end of each nine-week term. Support classes for special needs students were included, as long as they earned credit in one of the four core areas.

The data were compiled into a spreadsheet to calculate GPAs and PPAs as outlined above and perform statistical analyses. The results shown here are based on data collected through the cohort's sophomore year, generally including 15–18 graded terms.

Figure 2 compares the distribution of GPA and PPA values for the whole cohort using 0.5–point intervals. Clearly, calculating a measure of overall student proficiency that includes all individual classroom achievements counted on a linear scale (PPA) brings up the low end by "correcting" for the GPA calculation artifact that results from the scaling problem. For example, there are no students with a PPA below the 1.5–2.0 range and many more students have PPAs in the 3.0–4.0 range. If a college or university chose to use a lower limit of 3.5 for automatic admission about twice the number of students would qualify based on their PPAs than would qualify based on their GPAs.

Figure 2 compares the distribution of GPA and PPA values for the whole cohort... Clearly, calculating a measure of overall student proficiency that includes all individual classroom achievements counted on a linear scale (PPA) brings up the low end by "correcting" for the GPA calculation artifact that results from the scaling problem.

Figure 2. Distribution of GPA and PPA Values (Whole Cohort)



Because use of letter grades as a measure of student achievement is so engrained in our educational culture, it is important to note the following: GPA values are often "back-converted" into letter grades so that, for example, a student with a GPA of 2.0 is considered a C-average student (or C student), or a student with a GPA of 3.0 is considered a B student. Since calculation of a PPA does not involve letter grades it is incorrect to "back-convert" these values directly into letter grades. For example, a student with a PPA of 3.0 (on a four-point scale) performs on average at a 75 percent level, which corresponds to a C average (GPA around 2.0). This is also evident from comparing the graphs in Figure 1. A C student (GPA 2.0), for example, performs on average in the 70–80 percent range, which corresponds to PPA values between 2.8 and 3.2.

Table 1. Data Summary: Comparison of Mean GPA and PPA Values.[1]

Demographics	Count	Mean GPA	Std. Dev. GPA	Mean PPA	St. Dev. PPA	Diff. (GPA-PPA)	Effect size Dif/0.40
Whole Cohort	59	2.73	0.79	3.21	0.40	-0.48	-1.2
Male	30	2.64	0.77	3.18	0.36	-0.54	-1.4
Female	29	2.82	0.81	3.24	0.44	-0.42	-1.1
Minority[2]	13	2.42	0.98	3.05	0.49	-0.63	-1.6
White[3]	46	2.81	0.71	3.25	0.37	-0.44	-1.1
Special Ed.[4]	6	2.43	0.97	3.05	0.48	-0.62	-1.6
General Ed.[5]	53	2.76	0.77	3.23	0.39	-0.47	-1.2

[1] Data were statistically analyzed using paired t-tests. Levene's Test for Equality of Variances was not significant so equal variances were assumed (Levene's Test for minority versus white on the GPA was borderline significant [F=4.21, p=.05] but this does not affect any of the conclusions). Between males and females the GPA differences were not significant (t=.869, df=57, p=.389), and for the same two groups the PPA differences also were not significant (t=.485, df=57, p=.629. Between minority and white students both the GPA (t=1.629, df=57, p=.109) and PPA (t=1.564, df=57, p=.123) differences were not significant. The same was true between special ed. and general ed. students (on the GPA t=.957, df=57, p=.343; on the PPA t=1.032, df=57, p=.306).

[2] This category includes students self-identified as American Indian (2), Black (2), Hispanic (3), Asian (1), and Unspecified (5).

[3] Students self-identified as White.

[4] Students on an Individual Education Plan (IEP).

[5] Students not on an IEP.

Table 1 summarizes the means of the GPA and PPA values calculated for the whole cohort and broken down into different

The mean PPA values give a true measure of the difference in performance levels between the high and low achieving students, and again show that the low end is elevated, that is, the difference between relatively high and relatively low achieving students is not nearly as large as indicated by their mean GPAs.

demographic groups. The mean PPAs are notably higher than the GPAs both for the whole cohort and the various demographics. This is expected because for any classroom percentage below 90 percent the percentage point awarded (straight line in Figure 1) is progressively higher as the percentage decreases compared to the grade points awarded (discontinuous line in Figure 1).

For male vs. female (on average female students tend to outperform their male counterparts at least in the lower secondary grades), minority vs. white, and special education students vs. general education students, the differences between mean PPA and mean GPA are larger for the former than for the latter. Similarly, the differences in the mean GPA values between these groups are notably larger than the differences in the mean PPA values. The mean PPA values give a true measure of the difference in performance levels between the high and low achieving students, and again show that the low end is elevated, that is, the difference between relatively high and relatively low achieving students is not nearly as large as indicated by their mean GPAs. The statistical analysis (see footnote 1 to Table 1) reveals a critical fact; there are no differential effects between the various demographics.

To evaluate the magnitude of the differences between GPA and PPA for the whole cohort, as well as different demographics, effect sizes were calculated, as shown in the last column of Table 1. The effect size or standardized mean difference (Glass 1976) is useful to quantify the effect of a "treatment" when the means of a dependent variable are compared between a treated group and an untreated control group (see Marzano 2006). Effect size is defined as the difference between the means (treated minus control) divided by the standard deviation of the population. In our case the measure of overall student performance is the dependent variable. As we argue here, it is the PPA that should be used as the correct measure of over all student proficiency, and, therefore, we consider the traditional calculation of the GPA the "treatment" (or mistreatment in this case) so that PPAs represent the values for the control groups. The negative values for the effect sizes then indicate by how much the mean of each demographic is lowered by using the ordinal (GPA) grade point scale. To focus on the effects of differences in the means we have used the standard deviation of the whole cohort PPAs as the common divider. As this is somewhat arbitrary, it is not so much the absolute value of the effect size that is of interest here, although the effect sizes appear rather large, but the comparison of the effect sizes between the paired demographics, i.e., male-female, minority-white, and special-general education. In each case the difference between PPA and GPA and the effect size is larger for the traditionally underperforming group demonstrating in a different way that these students' GPAs are lowered disproportionately. PPAs thus provide a better estimate of the true magnitudes of the differences in achievement between various student demographics as measured by overall performance.

Conclusions

Common reactions from educators to the ideas and the results of the high school GPA project presented here generally have centered around two questions: "Why are we doing this?" (referring to the traditional GPA calculation)" and "Why didn't anyone think of this before?" (referring to the simple and logical alternative proposed here).

In answer to the first question we note that the letter (A-F) grading system has been around for almost a century since first proposed by Starch and Elliot (1912, 1913). Although it is hard to understand their claim that a letter grade system for in-class grading of assignments in itself would be less subjective than a point or percentage system, there is no question that letter grades have taken a firm hold on grading systems used in the US, in secondary and higher education, and has become the system of choice to determine and document student achievement in individual courses. As shown here, it is not documenting student performance by letter grades that is the problem, per se. Rather, it is the way grade points are assigned to letter grades and

then averaged that creates the problem, and is responsible for the GPA being an unreliable, unfair and imprecise tool for documenting overall student performance.

The practice of measuring student achievement in individual courses using letter grades to grade assignments suffers from the same main drawback as described here for the GPA calculation, i.e., the difficulty of averaging letter grades. Fortunately, with the advent of modern, computerized (Web-based) grading programs that automatically calculate classroom percentages with many options available to teachers, use of letter grades is unnecessary, and should be discouraged both in secondary and higher education. For teachers and instructors it is no more work submitting percentages rather than letter grades at the end of a grading period. The only requirement for employing a PPA-based system is that school districts and other institutions adopt student information systems that have the flexibility to calculate a variety of overall student achievement measures. There should be a transitional period in which GPA and PPA are used side-by-side so as to recalibrate admission and scholarship criteria.

In answer to the second question, a literature search yielded very little in terms of published work directly relevant to the question of alternatives to the GPA. Most interest in the published literature appears to be in evaluating the GPA as a criterion for admission to, or as a predictor for success, in the next level of education [see Bassiri and Schultz (2003); Vickers (2000); Attaway (1983); Hallock and Ommert (1997); Sadler and Tai (2007)]. Only one blog (http://sandwalk.blospot.com/2008/01/abolish-grade-point-average.html) expresses the wish to abolish letter grades and grade points but offers no alternative.

This study proposes that institutions that provide and/or use GPA as a measure of overall student achievement, including, public and private high schools, colleges and universities, and scholarship and grant agencies, seriously reconsider reliance on the traditional GPA, and consider changing to a measure of student performance as described here (PPA) that is directly based on, and proportional to cumulative classroom assessment. As we show here PPA presents a more nuanced, reliable, fair and direct measurement of overall student achievement, and also may provide a more realistic estimate of the magnitudes of the achievement gaps between various student populations. Most importantly, comparisons return PPA and GPA results showing no differential effects in important populations: special education/general education, minority/white, and male/female.

Acknowledgements: We thank our colleagues at the high school for making their classroom percentages available. We thank the principal of the high school, Dr. Bob Bolden, and the director of secondary services in the school district, Laurie Moses, for supporting this work. Special thanks go to the high school registrar, Jan Gordon, for meticulously maintaining the GPA project database.

JOHANNES J. VOLWERK was a science teacher at Sheldon High School in Eugene, Oregon (now retired).

GERALD TINDAL is a professor at the College of Education of the University of Oregon in Eugene.

REFERENCES

Attaway, A. E. (1983). *A comparison of four methods of calculating high school grade point average and rank in class (GPA).* Dissertation, University of Missouri-Columbia.

Bassiri, D., & Schultz, E. M. (2003). *Constructing a universal scale of high school course difficulty.* Journal of Educational Measurement, Vol. 40 (2), p. 147.

Fewster, S., & Macmillan, P. D. (2002). *School-based evidence for the validity of curriculum-based measurement of reading and writing.* Remedial and Special Education, 23(3), 149-156.

Glass, G. V. (1976). *Primary, secondary, and meta-analyses.* Educational Researcher, 5, 3-8.

Hallock, P. W., & Ommert, D. (1997). *The value added method of calculating grade point average.* NASSP Bulletin, Vol. 81 (585), p. 116.

Marzano, R. J. (2006). *Classroom assessment and grading that work.* ASCD, Alexandria, VA, p. 5-11.

Mosely, R. (2010). *Duck wins hike Kelly's earning.* Register Guard, February 6, 2010.

Sadler, P. M., & Tai, R. H. (2007). *Weighting for recognition: Accounting for advanced placement and honors courses when calculating high school grade point average.* NASSP Bulletin, Vol. 91(1), p. 5.

Starch, D., & Elliot, E. C. (1912). *Reliability of the grading of high school work in English.* School Review, 21, 442–457.

Starch, D., & Elliot, E. C. (1913). *Reliability of grading work in mathematics.* School Review, 22, 254–259.

Vickers, J. M. (2000). *Justice and truth in grades and their averages. R*esearch in Higher Education, Vol. 41(2), p. 141.