Equitable Use of Grades in Measuring Student Accomplishment:
Analyses and Recommendations of the Subcommittee on Grading

Report of the Educational Policy Committee
University of North Carolina at Chapel Hill

Report Prepared by the Subcommittee on Grading

Peter C. Gordon (Chair)
Andrew J. Perrin
Gwendolyn Sancar
Kevin G. Stewart

Analyses of grading patterns at Carolina and at other universities have shown that there is substantial variation in the grades assigned in different courses, variation that reflects differing grading practices across academic disciplines and instructors.[1] This report examines how variation in grading practices affects the validity of grade-point average (GPA) as an aggregate measure of student accomplishment. It shows that the Achievement Index[2], an alternative method of aggregating students' grades that takes into account differences in grading patterns across courses, provides a better measure of relative student accomplishment than does raw GPA. It recommends that undergraduate transcripts list student Achievement Index in addition to raw GPA and that University Distinction be awarded based on the Achievement Index. These recommendations are presented to Faculty Council in the form of proposed legislation at the end of this report.

GRADES AS A MEASURE FOR COMPARING STUDENT PERFORMANCE

One common use of grades is to compare the academic accomplishments of different students. Grades are used in decisions about admissions to professional schools and graduate programs, in employment decisions, and by the University in awarding distinction upon graduation. While this report recognizes that grades are used for purposes other than comparing students' performance, its analyses and recommendations proceed from the observation that the University systematically uses grades to compare students' performance. Because this is a necessary, consequential, and widespread use of grades, the University should present information about grades in such a way as to maximize the validity of the comparisons made between students.

Grade point average is a familiar measure of student performance that is commonly used in college and high school. It is the average of the numerical equivalents of a student's letter grades, weighted by the number of credits associated with each grade. GPA has long been recognized as a very problematic measure.[3] The most serious problem is that it is difficult to compare GPAs because students take different classes and the grading practices across classes vary substantially. The source of much of the variation between students' GPAs is therefore the courses and instructors students encountered, not the students' performance in those courses. Despite this problem, GPA is by far the most common aggregate measure used for the purpose of comparing the performance of different students.

The Achievement Index. The analyses and recommendations in this report focus on the Achievement Index (or AI), which was developed by Valen Johnson while he was in the Institute of Statistics and Decisions at Duke University, as an alternative to GPA that allows more valid comparisons of students' academic accomplishments. The AI is one of a number of statistical

---

[1] Quantitative analyses of grading patterns that document these patterns at Carolina can be found in two EPC reports: *Grade Inflation at UNC Chapel Hill (*http://www.unc.edu/faculty/faccoun/reports/R2000EPCGrdInfl.PDF) and the *Annual Report for 2004* (http://www.unc.edu/faculty/faccoun/reports/2003-04/R04EDP1.htm). Other recent discussions of grading at Carolina can be found in the 2001 *Report of the Task Force on Grading Standards* (http://www.unc.edu/faculty/faccoun/reports/2000-01/R2001GradingStandards.htm) and *EPC's Annual Report for 2005* (http://www.unc.edu/faculty/faccoun/reports/2004-05/R05EPC1.htm).

[2] Johnson, V.E. (1997). An alternative to traditional GPA for evaluating student performance. *Statistical Science, 12*, 257-278; Johnson, V.E. (2003). *Grade inflation: A crisis in college education*. New York, NY: Springer.

[3] In addition to the works by V. Johnson, see Stricker, L.J. et al. (1992). *Adjusting college grade point average for variations in grading standards*, (Research Report, Educational Testing Service) and Lei, P.W. et al. (2001). *Alternatives to the grade-point average as a measure of academic achievement in college*, (ACT Research Report Series).

methods developed to address problems with GPA. Other methods which have been subjected to serious analysis include procedures based on linear regression and on Item Response Theory (to which AI is related).[4] The AI is used here because of the thoroughness with which it has been examined and because its assumptions about the meaning of grades seem to be the most appropriate for college grading.[5] Further, the AI builds on the current system of grading and places no restrictions on faculty members' use of specific grades, a characteristic that is advantageous in practice and also in principle because it recognizes faculty members' academic freedom.

The Achievement Index (AI) is a measure of a student's performance relative to all other students taking classes at the institution at which the student is taking classes. While calculation of the AI is complex, its underlying principles are simple and relate in a straightforward way to issues in college grading. A brief summary of the AI is given below.

- Assumptions about the meaning of grades. The AI makes the minimal assumption that in all courses grades provide ordinal information about the performance of students in a class; grades are not assumed to provide further information (particularly about absolute levels of performance) in a way that can be combined meaningfully across courses. Building on this minimal assumption, the grades assigned in a class are used to categorize students into groups receiving the same grade and to rank those groups relative to other groups (e.g., students getting A are in the top-ranked group, students getting A- are in the second-ranked group, etc.). The number of ranked groups in a class depends on the number of different grade categories that are used by the instructor. For example, there would be three ranked groups in a class where the only grades used were A, A- and B+ while there would be four ranked groups in a class which used the grades A-, B+, B and B-. In this example an A- in the second class would be equivalent to an A in the first class because both indicate placement in the top-ranked group.

- Using the ordinal information provided by grades. A student's AI is a summary score indicating the student's academic performance relative to the performance of other students taking classes at the University; it is calculated by determining what score is most likely for the student given the ordinal information about grades in the courses that the student took. This calculation takes into account the fact that students take different courses from different instructors, and that they do so with different classmates. Combining information across these different classes requires information about the criteria that individual instructors use to assign grades; criteria that are expressed as cutoffs for the percentage of students that are assigned to different grade categories (top category, next best category, etc.). Those cutoffs must be estimated with respect to all students, not just the students in a given instructor's class. Therefore, the AI calculation evaluates different cutoffs for grade categories in a class until it finds the cutoffs under which the grades in the class are most likely to occur given the AI scores of the students in the class. Information about the cutoffs for classes is then used to refine the students' AI scores, which in turn are used to obtain a better estimate of

[4] Larkey, P. & Caulkin, J. (1992). *Incentives to fail*. Working Paper 92-51, Heinz School of Public Policy and Management, Carnegie Mellon University. Young, J.W. (1990). Adjusting the cumulative GPA based on item response theory. *Journal of Educational Measurement, 12*, 175-186.
[5] The EPC Annual Report in 2005 discusses the merits of the AI in relation to approaches for adjusting GPA based on linear regression and also in relation to the procedure commonly used in high schools of assigning higher numerical values to grades in AP and honors classes as compared to regular classes.

instructors' grade cutoffs. This back and forth process is repeated until the set of student AI scores and the set of instructor grade cutoffs are found for which the observed pattern of all students' grades in all classes is most likely to occur.

- Consequence 1. The grade that a student receives in a class contributes to that student's AI only to the extent that it provides information about that student's achievement relative to other students' achievement. Thus, in a class where all students get the same grade (e.g., an A) that grade neither helps nor hurts the student's AI.

- Consequence 2. The meaning of the ordinal information provided by the ranked grade categories in a class depends on the overall academic performance of the students taking the class (e.g., it is a greater achievement to outperform students who generally do well in their other classes than it is to outperform students who generally do poorly in their other classes).

Assessment of AI using grades at Carolina. For this assessment, GPA and AI were computed for students earning a Bachelors degree at Carolina from May of 1999 through May of 2004.[6] While GPA and AI are different ways of looking at student accomplishment, it is noteworthy that the views that they provide are far more similar than they are different, with the rankings of student accomplishment as measured by GPA and by AI having a correlation of .926 (a correlation of 1.0 would mean that the two rankings were identical). This comes as no surprise since GPA and AI are computed from the same grades. While there is a great deal of similarity between the results obtained by the two measures there are differences as well. In cases where they differ, the validity of the rankings based on GPA and AI can be assessed by examining whether GPA ranking or AI ranking is a better predictor of performance by students enrolled in the same class at the same time. This is done on the assumption that instructors apply the same standard to all students in a class and that when evaluated by the same standard more accomplished students tend to do better than less accomplished students. This analysis was done for all pairs of students where one student in the pair was ranked higher by GPA but the other student was ranked higher by AI. In 61% of cases where students in such pairs earned different grades in the same class, the higher grade was earned by the student who ranked higher on AI but lower on GPA. This shows that where GPA and AI diverge, AI provides a more valid measure of student performance.[7]

---

[6] The analyses presented in this report were based on a data set, compiled by the Registrar's office, containing all grades given in courses taken by undergraduates from Fall 1995 through Spring 2004. Results are presented beginning with the graduating class of Spring 1999 because most students graduating earlier had taken many courses before the period covered by the data set. Further, the analyses did not include students who took 10 or fewer classes during the time of the analyses. Exploratory analyses showed that the pattern of results was not influenced much by the number-of-classes criterion for inclusion. Finally, the analyses did not include classes with fewer than five students because calculating AI in those cases requires assumptions that are not necessary for larger classes. Calculation of the AI was performed by Valen Johnson (now at the M.D. Anderson Cancer Center of the University of Texas), whose assistance is gratefully acknowledged.

[7] This method assesses what is called internal order consistency. The ordering of students by aggregate measures (in this case GPA and AI) is evaluated using the ordering of students on that subset of measures (in this case classes) that allow students to be compared directly. Another method for assessing the relative validity of different aggregate measures is to compare the strengths of their relationships to a measure of student performance that is external to college grading. Unfortunately, no external measures of this sort are generally available for college students while they are in college or after they graduate. The only available external measures reflect student performance prior to college and consist of performance in high school and on standardized tests. The multiple correlation of GPA with high-school class rank and SAT scores in the group studied here was .516 (26.6% of the variance) and for AI it was .595 (35.4% of the variance). While this pattern supports the idea that AI is a better

The question of whether there are consequences associated with the divergences between AI and GPA was examined by looking at the awarding of different levels of distinction (Distinction and Highest Distinction), which at Carolina are awarded on the basis of student GPA. To the extent that GPA measures characteristics of courses and instructors rather than student performance, using GPA is an inaccurate and unfair way to award distinction.   During the period studied for this report, the top 5.6% of students graduated with highest distinction (GPA of 3.8 or higher) and the next 17% graduated with distinction (GPA of 3.5 to 3.8).[8] Those percentages were applied to students ranked by AI in order to determine an alternative AI-Distinction which could be compared to the currently used GPA-Distinction. Table 1 shows the numbers of students receiving distinction under the two methods. Approximately 26% of students merit distinction by at least one of the methods while the remaining 74% of students do not merit distinction by either method. Of the students who merit distinction by one of the methods, approximately 64% are categorized in the same way (Distinction or Highest Distinction) by both methods. The remaining 36% of those students receive different levels of distinction depending on which method is used (GPA-Distinction or AI-Distinction), which shows that the proportion of students who switch distinction categories is large. This indicates that variation in grading across courses, instructors, and departments, which are taken into account by the AI but not by GPA, exerts considerable impact on which students receive distinction at Carolina.

<div align="center">

Achievement Index

		None	Distinction	Highest
	None	15667	769	1
GPA	Distinction	766	2609	244
	Highest	4	241	945

</div>

Table 1. Number of students receiving different levels of distinction using the current GPA method and the AI method. Students on the diagonal (numbers with background shading) receive the same level of distinction by GPA and AI. Students above the diagonal (numbers surrounded by a dark line) get higher distinction by AI than by GPA. Students below the diagonal (numbers surrounded by a double line) get higher distinction by GPA than by AI.

The validity of the distinction categories created from GPA and AI were assessed by examining whether GPA-based distinction or AI-based distinction was a better predictor of performance by students enrolled in the same class at the same time. There were 22,296 cases where the students in such a pair earned different grades in the same class. In 14,455 of those cases (65%) the student with AI-only Distinction earned a higher grade than the student with GPA-only Distinction. For Highest Distinction, there were 1484 relevant pairings and for 914 of them (62%) the student with AI-only Highest Distinction earned a higher grade than the student with GPA-only Highest Distinction. Thus, when taking exactly the same class students in the

---

aggregate measure than GPA, this report focuses on internal consistency measures because they rely solely on evaluations made at Carolina and make no assumptions about the relevance or validity of high school class rank or SAT scores.

[8] Determination of distinction was based on GPA calculated from grades in the data set, not from actual award of distinction, because in some cases grades that were not in the data set contributed to the GPAs used for determining distinction. Exploratory analyses showed that the pattern of results was not influenced much by using actual distinction rather than calculated distinction.

AI-only groups, who currently receive no or lesser distinction, got a higher grade almost twice as often than did students from the GPA-only groups, who currently receive higher distinction.

RECOMMENDATIONS

The analyses presented above show that AI provides a more accurate (and hence more equitable) basis for comparing student performance than does GPA. Accordingly it is recommended that Carolina should:

- **List AI on students' transcripts**. AI should be calculated from students' course grades and listed on their transcripts in addition to GPA. AI scores are real numbers distributed around zero. However, a simple rescaling procedure transforms those scores onto the familiar four-point scale used for grades. This rescaled AI would be used on students' transcripts and would be described as either Achievement GPA or Adjusted GPA.

- **Award distinction based on AI**. The current system for awarding distinction was adopted by Faculty Council in the mid 1980s, and while it is based on absolute levels of GPA one consideration in choosing was that about 5% of students would be expected to earn Highest Distinction and an additional 10% would earn Distinction given the grade levels at the time. In Spring of 2004, 5.7% of graduating students received Highest Distinction and 16.9% received Distinction. Using AI for awarding distinction requires re-considering the percentage of students whose performance should be recognized with different levels of distinction. The selection of percentage cutoffs to be applied to the AI in awarding distinction need not be governed by the considerations used in the mid 1980s. In addition, it might be desirable to change from two to three levels of distinction (Distinction, High Distinction and Highest Distinction).

- **Determine class rank using AI**. Upon request, the Registrar's Office provides students with a letter indicating their class rank based on GPA. Instead, class rank would be determined using AI.

Additional review will be needed to determine whether there are other cases where the University currently uses GPA but instead should use AI. For example, at present GPA is used at Carolina for admission to undergraduate professional schools. A decision about whether to use AI rather than GPA should probably be left to the schools that make the admissions decisions, though it might be beneficial for the University to formulate guidelines about how AI could be used. As a second example, the criteria for awarding Departmental Honors at Carolina are formulated by departments, which do so within the framework of University guidelines that include GPA; again, consideration would need to be given to the question of whether those guidelines should be reformulated. Finally, GPA plays a role in standards for graduation and continuing academic eligibility. However, graduation and continuing eligibility should not be based on comparative standards and therefore AI should not be incorporated into those standards.

Implementation. A number of important issues would have to be addressed in order for the AI to be used at Carolina in the manner described above. We recommend that an implementation committee composed of faculty and administrators be constituted and charged with resolving technical and procedural issues associated with using the AI. The critical technical issues center on methods for secure and reliable calculation of the AI. Critical procedural issues include questions such as how to define cohorts of students for purposes of determining distinction and class rank, and how to communicate information about the AI.

GPA may be a misleading measure but it is very familiar. AI is not familiar, so procedures would have to be developed for communicating information about AI both internally to students and teachers at Carolina and externally to employers and other educational institutions that consider the academic performance of our students. Procedures for internal communication would include informing students about their AIs after every semester in which they take classes and also providing information to students and teachers about the extent to which classes could impact the AI of an enrolling student based on the degree to which the classes have differentiated levels of student performance in the past. This latter type of information would complement the information that is currently available to students from *Pick-a-Prof*, a commercial service that provides students with information about the grades in individual sections of courses (listed by instructor) at Carolina and at other public universities. Procedures for communicating externally about the AI would have to be developed at a number of levels, ranging from a simple one-sentence statement about the AI, which would appear on documents such as transcripts, to more comprehensive discussions of the mechanics of the AI and the reasons that the University has adopted it.

It is also recommended that during the period in which use of the AI is being implemented, students and faculty should be given information about what the AI would say about student performance and about how classes contribute to students' AIs. While that information would not be part of a student's official record, it would help make the concept of the AI familiar and would also allow the AI implementation group to get feedback on different options for presenting information. Finally, it is recommended that procedures be put in place for tracking the impacts of the AI.

Resolution 2007-__. On Adopting the Achievement Index as the Metric for University-wide Comparative Rankings of Students.

Whereas, interdepartmental and inter-instructor variation in grading in undergraduate courses has been identified by the Educational Policy Committee as a serious and ongoing concern; and

Whereas, although departmental and disciplinary grading standards may appropriately vary in consequence of the philosophies and orientations of the disciplines, nevertheless such grading disparities constitute a specific threat to the validity of University-wide comparative rankings of undergraduate students based solely on grades; now therefore

The Faculty Council resolves:

Pursuant to the recommendation of the Educational Policy Committee in its report to the Council of March, 2007, the Achievement Index [as defined in Valen E. Johnson, "An Alternative to Traditional GPA for Evaluating Student Performance," *Statistical Science*, Vol. 12, No. 4 (Nov. 1997), pp. 251-269] is adopted as the metric for University-wide comparative rankings of undergraduate students, including but not limited to the awarding of University distinction.

The University Registrar is requested to record the student's Achievement Index (AI), or a derivative measure, on the official undergraduate transcript in addition to the traditional Grade Point Average (GPA). The Registrar shall also provide appropriate documentation to allow those relying on transcript information to interpret the AI.

The Provost is requested to appoint an AI Implementation Task Force charged to make recommendations concerning the institutional and technical steps needed to insure the long-term viability of, and stakeholders' confidence in, the calculation, recording, and dissemination of the AI. The task force is also charged with developing procedures for tracking the impact over time of these changes.