The Achievement Index (AI) was developed by Valen Johnson as a method for combining the information from grades earned in different college classes; a complete presentation of the index and how it is calculated can be found in Johnson (1997)[1], with additional discussion available in Johnson (2003)[2]. The overarching goal of the AI is to measure each student's academic performance while factoring out differences among individual instructors' grading practices.

The principles underlying the AI have substantial overlap with a set of methods known as Item Response Theory or IRT (see Embretson & Reise, 2000 for a review)[3]. IRT is the statistical foundation for almost all standardized tests of academic aptitude or achievement (it is used in the SAT, ACT and GRE; in North Carolina it is used for the end-of-grade and end-of-course tests that are given in the public schools). IRT provides a set of methods for jointly estimating the ability of test takers, the difficulty of test items, and the precision of a test across the range of abilities in the population of test takers. IRT methods are used to score students' test performance, develop test items, and maintain comparability of test scores across different versions of a test. While IRT was originally developed for test items that have a binary scoring (e.g., correct/incorrect), polytomous (or graded) models extend IRT to cases where items are scored using an ordered series of categories (e.g., poor, fair, good and excellent). The AI is a further application of polytomous IRT models. The classes that a student takes are considered analogous to items on a test and grades in classes constitute the ordered series of evaluative categories; crucially, the AI extends some components of IRT to the situation where the test takers select the test items that they will attempt to answer (i.e., students select the classes that they take). Thus, the AI allows use of some of the powerful statistical methods of standardized testing without having to standardize the knowledge being evaluated or using a limited set of easily-scored test formats.

The basic idea of the AI is to measure each student's academic performance using comparisons of how that student did relative to his or her classmates. Grades within a class are assumed to provide ordinal information about the categories of student performance (i.e., an A is assumed to be better than a B, but no further assumptions are made about how much better one grade is than another). Figure 1 illustrates how student performance is grouped by the AI for four sections of the same course. The upper bar in the chart for each section shows the percentage of students receiving various letter grades from the instructor. For example, in Section A there were 5 As (22.7%), 5 A-s (22.7%), 5 B+s (22.7%), 4 Bs (18.2%) and 3 B-s (13.6%). Because students take courses from different instructors and do so with different classmates, the AI calculation determines the cutoffs used for assigning grades in each class. Those cutoffs are characterized in terms of what level of performance, across the range of academic performance by all students taking classes at the institution, must be achieved in order to earn a specific grade. In the charts in Figure 1, the lower bar for each section shows the cutoffs for the different grades given in that section. For Section A, the grade of A implies that a student is above the 61st percentile of the student body; an A- implies that a student is above the 32nd percentile and below the 61st percentile; a B+ implies that a student is above the 17th

[1] Johnson, V.E. (1997). An alternative to traditional GPA for evaluating student performance. *Statistical Science, 12*, 257-278.
[2] Johnson, V.E. (2003). *Grade inflation: A crisis in college education*. New York, NY: Springer.
[3] Embretson, S.E., & Reise, S.P. (2000). Item response theory for psychologists. Mahwah, NJ: Erlbaum Associates.

percentile and below the 32^{nd} percentile; a B implies that the student is above the 8^{th} percentile and below the 17^{th} percentile; and a B- implies that a student is below the 8^{th} percentile.

Calculation of the AI for a student is based on mapping the letter grades the student received in her classes onto the percentile ranges in the population of students that are associated with those grades. Accordingly, a student's AI is likely to be raised more by an A in Section B, where getting an A implies exceeding the 88^{th} percentile in the student body, than by an A in Section A, where getting an A implies exceeding the 61^{st} percentile. In turn, calculation of the grade cutoffs for a class is based on the academic performance of the students receiving various grades as indicated by the AIs of those students based on their grades in other classes.

A full mathematical description of how the AI is calculated is presented in Johnson (1997). Here we give a schematic outline, adapted from Johnson (2003), of the calculation.

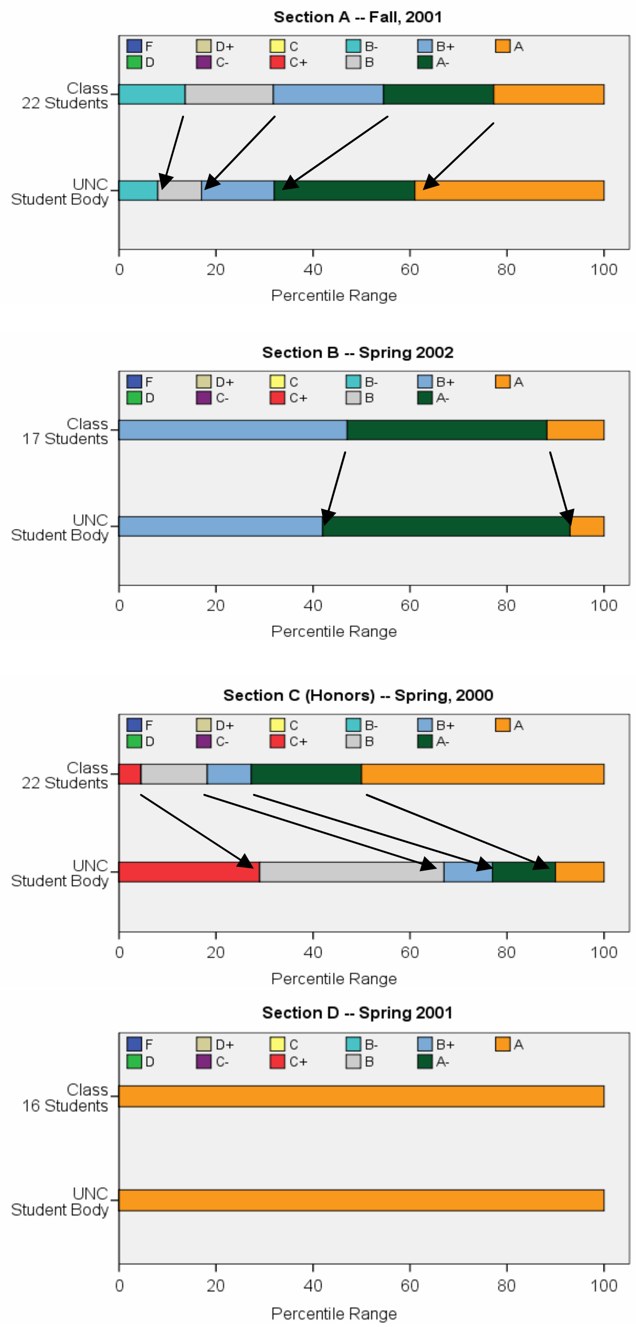

Figure 1. Results of applying the AI to four sections of the same course. The top bar for each section indicates the percentage of students receiving various grades from the instructor. The bottom bar shows corresponding percentile ranges of those grades in terms of academic achievement in the entire student body.

The overall academic performance of every student is assumed to vary along a normal distribution, with the AI indicating the mean of that distribution. The AI is determined by finding the value of the mean that best matches the student's grades in courses. Figure 2 illustrates how this is done for an example where a student has taken three courses, getting a B in Course 1, an A- in Course 2, and a B+ in Course 3. Scenario 1 demonstrates the match between those grades and an AI of 3.3. The quality of that match is given by the product of the area under the normal distribution of the students performance that falls within the grade cutoffs for the grade that the student received in each of the three courses (as indicated by the shaded areas for the three curves in Scenario 1). Scenario 2 shows the match between the same three grades and an AI of 3.8 (this higher mean shifts the distribution to the right). The product of the shaded areas in that instance is less than in Scenario 1, therefore Scenario 1's AI of

3.3 is a better match for the student's grades. Calculating the AI involves finding the mean (position) of the distribution that maximizes the product of the shaded areas.

As illustrated in Figure 3, calculation of the grade cutoffs used in courses proceeds in a similar way. Here, the means (positions) of the distributions representing the AIs of the students in the class are held constant and the position of the grade cutoffs are adjusted until the product of the shaded areas under the curves is maximized.

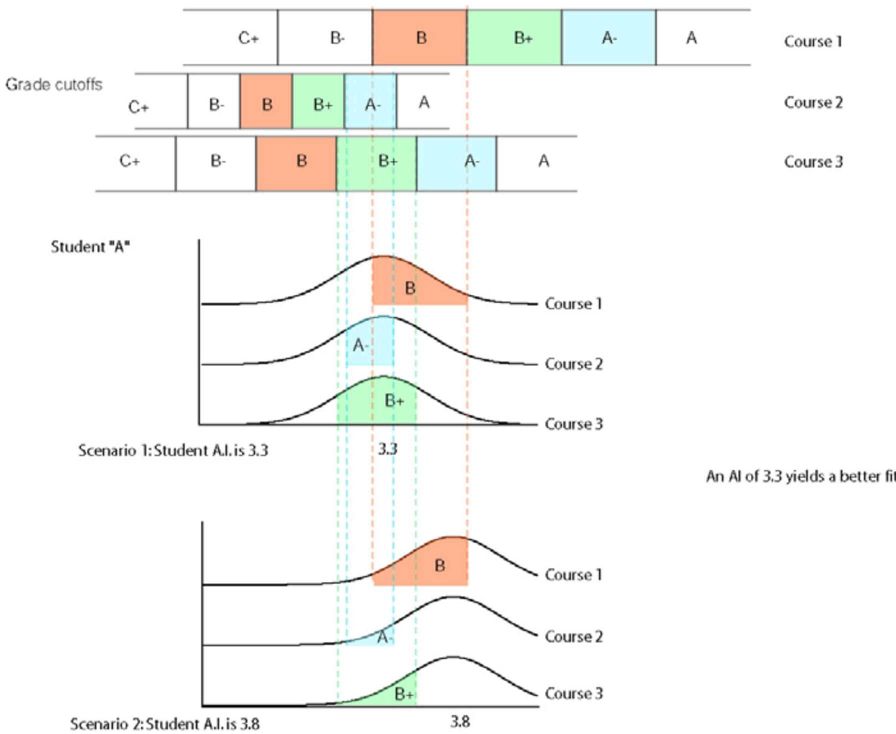

Figure 2. Illustration of how a student's AI is calculated.

Final calculation of the set of student AIs and the set of class grade cutoffs is achieved through an iterative process where student AIs are calculated and then used to refine calculation of class grade cutoffs, which in turn are used to refine calculation of student AIs. This process continues until the sets of student AIs and grade cutoffs are found that best match the grades all students received in all classes. The manner in which the AI is calculated has two very important consequences.

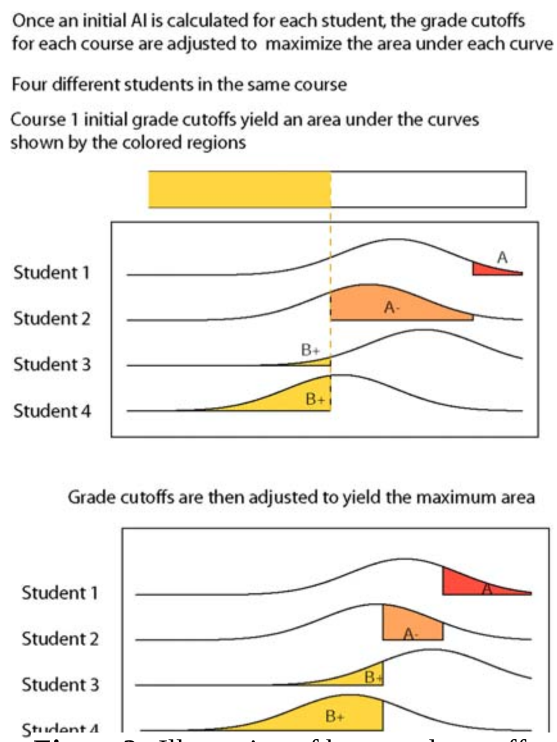

Figure 3. Illustration of how grade cutoffs are calculated.

First, because grades within a class only have ordinal properties, the impact of those grades on students' AIs depends on the extent to which those grades differentiate levels of student academic performance. For example, in Section B (shown in Figure 1), only a small proportion of students receive the grade of A, which magnifies the value of those As as compared to Section A where a higher proportion of As are awarded. In contrast, relatively large proportions of students in Section B receive A- or B+ and those grades are associated with broad ranges of student achievement. In Section D, where every student received an A, the instructor's grading provided no differentiation of student performance, so the grade of A only implies that a student exceeded the 0^{th} percentile. In other words, the probability of a student getting the grade of A is 1 regardless of the mean (position) of the student's AI. Because a student's AI is set to the value that maximizes the product of the probability of the grades that are observed in all the student's courses, the grade in this class has no effect—positive or negative—on the student's AI.

Second, the impact of grades in a class depends on the academic achievement that the students in the class have demonstrated in their other classes. In Section C, 50% of the students received an A, yet receiving an A in that class implies that a student exceeds the 90^{th} percentile in the student body. This occurs because Section C is an Honors section where many of the students have demonstrated relatively high academic achievement in their other courses; this has a strong effect on the grade cutoffs for the class.

Figure 4 presents both GPA and AI information for transcripts from two students who graduated from Carolina with the same major. Transcript 1 shows a higher GPA than Transcript 2, while Transcript 2 shows a higher AI than Transcript 1. The students' grades in individual classes are listed along with the percentile ranges for different grades in the class in which the grade was assigned. For both students, the percentile ranges associated with different grades varies a great deal from class to class. However, there is a strong tendency for high grades to include lower percentiles of student performance in the classes listed in Transcript 1 as compared to Transcript 2; this is especially so for classes in which Transcript 1 shows high grades. This difference in courses plays a role in calculating the AI but not in calculating GPA.

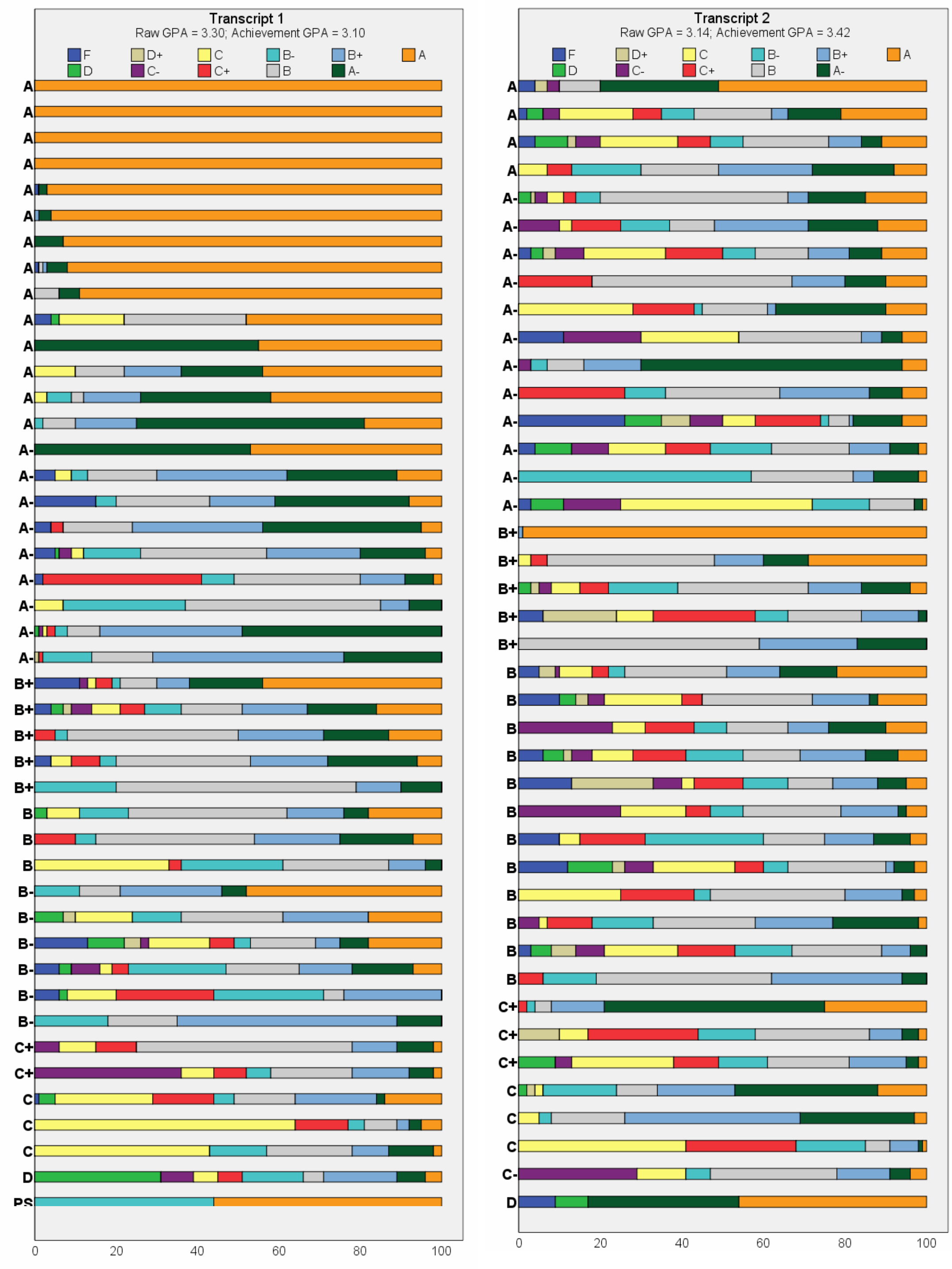

Figure 4. Sample transcripts of two students with the same major. Grades in classes, sorted from highest to lowest, are listed on the left-hand side for each student. The regions of the adjacent bars show the percentile ranges for different grades in each of the classes that taken by the students.