NCME Instructional Module on

Obtaining Intended Weights When Combining Students' Scores

Albert C. Oosterhof
Florida State University

An instructor typically combines students' scores from several measures such as assignments and exams when assigning course grades. The relative weights intended for these scores are at least inferred and often stated explicitly by the instructor. This module describes how scores can be adjusted so that the intended weights are obtained. Techniques are discussed for two grading criteria: (a) grading students through comparison to others in the class and (b) grading students through comparison to predetermined levels of performance.

When an instructor bases course grades on more than a single assignment or test score, intended weights of these scores are often communicated explicitly to the students. For example, an instructor might tell students that each of the five assignments represents 10% of the grade and each of the two tests represents 25% of the grade. On the other hand, an instructor might choose not to apply any particular weights when combining these scores into a single grade. In either case, the mathematical principles involved when combining scores ultimately determines the relative weights that each of the scores will receive. These principles are not complicated to apply. But unless the instructor incorporates these principles when combining scores, it is likely the relative weights realized by scores will be different from those that the instructor announced to students, or at least inconsistent with those the instructor would choose to defend.

Many instructors control the weights of scores by altering the number of points associated with each test and assignment. If students can earn twice as many points on one test as on another, that test is assumed to have twice the weight. In contrast, textbooks that discuss grading often indicate that the *variability* of scores determines their weight. If scores on one test spread out twice as far from the average score as scores on another, the former test is expected to have twice the weight of the latter.

This paper first describes the effects that maximum points and variability of scores have on the weights of scores. From this information, procedures are described that help obtain intended weights when combining students' scores into course grades. It will be demonstrated that using inappropriate weighting procedures can significantly affect the grades assigned to students. This paper is designed to help you achieve three skills: (a) determining when the maximum number of obtainable points affects the weight of scores, (b) determining when the variability of scores affects their weight, and (c) using the appropriate procedure for weighting scores when assigning course grades.

Albert C. Oosterhof is a professor in the Department of Educational Research at Florida State University, Tallahassee, FL 32306. His specializations are applications of measurement in the classroom and computer applications in testing.

Series Information

ITEMS is a series of units designed to facilitate instruction in educational measurement. These units are published by the National Council on Measurement in Education. This module may be photocopied without permission if reproduced in its entirety and used for instructional purposes. Barbara S. Plake, University of Nebraska–Lincoln, has served as the editor for this module.

Skill 1: Determining When Maximum Number of Points Affects the Weight of Scores

When the percentage of points earned is the basis for assigning grades, the weight of each score is affected by the maximum points a student can earn on that score, as demonstrated below (Example 1).

	Exam Scores		Total	Total
	1st	2nd	points	percentages
James	0	100	100	83
Laura	18	90	108	90
Tony	20	80	100	83
Maximum obtainable	20	100	120	

Each of three students has scores on two exams. The maximum number of points that can be obtained on the second exam is five times that of the first. When each student's scores are totaled, the second exam influences the total percentage of points more than the first. James earned none of the possible 20 points on the first score, and all of the 100 points on the second score. If these scores were weighted equally, James would have earned a total score midway between 0% and 100%, or 50%. Instead, his total score is 100 of a possible 120 points, or approximately 83%. This total percentage is five times closer to the second score than the first. Tony's total percentage is similarly affected five times as much by his second than by his first score. He obtained 20 of the 20 points on the first score, and 80 of the 100 points on the second. If the scores were weighted equally, Tony would have earned a proportion of points halfway between 100% and 80%. However, his total score is 100 of 120 points, or approximately 83%. Even though James earned zero points on the first exam, his total percentage score is the same as Tony's. This is because James earned 100% of the points on the more heavily weighted second exam. Laura consistently received 90% of the possible points across the two exams; therefore, the relative weights of her two scores had no effect on the percentage of total points she earned.

If course grades are based on percentage of total points, and two assignments or exams are to receive the same weight, the maximum points obtainable on each score must be made equal.[1] A convenient procedure for equating the maximum points is to convert all scores to percentage scores. Step 2 of Table 1 illustrates this approach. By converting scores to percentages, the weights of the two exams have become equal. Total percentages obtained in Step 2 by James, Laura, and Tony are midway between the percentage of points each earned on the first and second exams.

If the weights of scores are first equated, intended weights are obtained by multiplying the equated scores by the desired weights. If an instructor wishes to give a first exam twice the weight of the second, each student's score on the first exam is multiplied by two. This is illustrated in Step 3 of Table 1.

Only the maximum obtainable score affects the relative weight of each score when grades are based on percentage of total points earned. If you intend for one assignment or test to have twice the weight of another, simply associate twice the points with that score. Verify your understanding of this particular concept by evaluating the four approaches presented below for weighting scores. Assume that, within each situation, grades are to be based on scores from two assignments and one exam. Each assignment is to represent 25% of the grade, and the exam is to represent the remaining 50%. Does each approach give the intended weights?

Approach 1. The maximum score on each of the two assignments is set at 25 points, and the maximum score on the exam is set at 50 points.

Approach 2. The maximum score on each of the two assignments is set at 10 points, and the maximum score on the exam is set at 20 points.

Approach 3. The maximum scores on each of the two assignments and on the exam are set at 20 points. However, each student's score on the exam is doubled before scores are totaled for each student.

Approach 4. The maximum score is set at 10 points on the first assignment, 20 points on the second assignment, and 60 points on the exam. The scores on the first assignment are doubled to bring the total maximum score to 100.

Answers. Approach 1 will result in the intended weights because the maximum number of points associated with the three sets of scores is proportional to the intended weights. Approach 2 will result in the intended weights for the same reason as Approach 1. The maximum number of points associated with the exam is twice that of each of the assignments. In Approach 3, the three sets of scores started out with equivalent weights because the maximum number of points a student could earn within each set was the same. However, because the exam scores are doubled before the scores are totaled, the exam obtains the intended weight. Approach 4 results in maximum scores of 20, 20, and 60. Therefore, the exam has three times the weight of each assignment. The fact that maximum scores on the assignments and the exams now total 100 points is irrelevant to the solution of the weighting problem. The intended weighting was not realized.

Skill 2: Determining When the Variability of Scores Affects Their Weight

When grades are based on the percentage of points earned, the variability of scores has *no effect* on the weight of the scores, as illustrated below (Example 2).

	Assignment 1st	Assignment 2nd	Total points	Total percentages
Cheryl	24	24	48	96
Kimberly	19	25	44	88
Robyn	14	23	37	74
Maximum obtainable	25	25	50	
Range	10	2		

The maximum a student can obtain on each assignment is 25 points, although the range of scores on the first assignment is five times greater than that of the second. Because the weights of scores are determined by the maximum points a student can obtain but not the variability of scores, scores on both assignments have equal weight. Each assignment equally affects the *total percentage of points* obtained by students. For example, Kimberly received 19 of 25 points, or 76%, on the first assignment and 25 of 25 points, or 100%, on the second. The total percentage of points she obtained

TABLE 1

Procedure for Obtaining Intended Weights When Grades Are Based on Percentage of Points

	Step 1: Exam scores		Step 2: Percentage scores				Step 3: Weighted scores			
	1st	2nd	1st	2nd	Total	%	1st	2nd	Total	%
James	0	100	0	100	100	50%	0	100	100	33%
Laura	18	90	90	90	180	90%	180	90	270	90%
Tony	20	80	100	80	180	90%	200	80	280	93%
Maximum obtainable	20	100	100	100	200		200	100	300	
Relative weights			1	1			2	1		

is 88%, midway between the percentage of points she earned on the two assignments.

However, when grades are based on a student's performance relative to other students, the weight of scores *is* affected by the variability of scores. An assignment or exam on which students obtain similar scores ends up with less weight when scores are totaled. This is illustrated below (Example 3).[2]

	Exam Scores 1st	2nd	Total score
Angela	38	90	128
Melvin	42	88	130
Vicki	46	86	132
Maximum obtainable	50	100	
Average score	42	88	
Range of scores	8	4	

Among the three students, Angela obtained the lowest score on the first exam, Vicki obtained the highest score, and Melvin obtained a score midway between the other two scores. On the second exam, the pattern is reversed, with Angela receiving the highest score, Vicki the lowest score, and Melvin again receiving a score midway between the other two scores. Given that the relative performance of the three students on the first exam is exactly opposite their performance on the second, had the first and second scores been weighted equally, the *relative standing* of all three students would be equal. In reality, the ranking of students' total scores is affected more by the first than by the second exam. Angela obtained the lowest total score, and Vicki obtained the highest total score. Scores on the first exam actually have twice the weight in determining the eventual ranking of students because scores on that exam have twice the variability (e.g., the range from highest to lowest scores is twice as large on the first exam).

When grades are assigned by comparing students to other students, a high grade is not assigned as a result of earning a high percentage of points. Instead, a student receives a high grade when he or she earns *more* points than other students. In this context, the second exam in Example 3 has less weight in determining students' rankings even though twice as many points are associated with that exam and its average score is more than twice as large. When grades are assigned by comparing students to other students, the weights of assignments and exams are affected and consequently controlled by the variability of their scores.

"Standard Deviation"—A Preferred Measure of Score Variability

The variability of scores has been shown to affect the weight of scores when students are graded through comparisons to others. To determine the weight of each set of scores, and to eventually alter these weights to desired values, we must establish an *index of score variability.*

In Example 3, *range* is used as the index of variability. The range of scores on the first exam is 46–38, or 8 points. The range of scores on the second exam is 4 points. Therefore, the first exam is said to have a greater variability of scores and thus the greater weight. Unfortunately, range is an unstable measure of variability because it depends on the values of only the highest and lowest scores. Range does not take into account the variability among the scores between these two extreme scores. For example, if in a class of 30 students the student receiving the highest test score obtained five additional points, the range of scores would increase by five points even if the variability of all other scores remained unchanged. A preferable index of score divergence reflects the variability among *all* students' scores.

"Standard deviation" is a preferred index of variability. In fact, when assigning grades through comparison of students to each other, *the weight of one set of scores relative to other sets of scores is proportional to the standard deviations of the respective sets of scores.*[3] Step 1 of Table 2 redisplays the exam scores presented in Example 3; however, the standard deviations of these exam scores are now provided. Because the standard deviation of the first scores (4.0) is twice that of the second scores (2.0), the weight of the first scores is twice that of the second scores.

Procedures for calculating standard deviation are discussed in many introductory statistics and educational measurement textbooks. Here, focus is on a characteristic of standard deviation that allows us to obtain desired weights for scores.

The numerical value of standard deviation changes as the variability of scores changes. Sets of scores which are more spread out have a larger standard deviation. In Step 1 of Example 3, the scores 38, 42, and 46 on the first exam are twice as spread out among themselves as the scores 90, 88, and 86. Consequently, the standard deviation of the first scores is twice that of the second scores (4.0 vs. 2.0). Note that the numerical value of standard deviation has nothing to do with the average value of the scores, but pertains only to how much the scores spread out from each other.

Let us use this information to generalize to other scores. Given that the standard deviation of the scores in Set 1a of the illustration below (Example 4) is 4.00, estimate the standard deviations of the scores in Sets 1b, 1c, and 1d. Likewise,

TABLE 2

Procedure for Obtaining Intended Weights When Grades Are Based on Students' Relative Standing

	Step 1: Exam scores 1st	2nd	Step 2: Equated scores 1st	2nd	Total	Step 3: Weighted scores 1st	2nd	Total
Angela	38	90	9.5	45	54.5	9.5	90	99.5
Melvin	42	88	10.5	44	54.5	10.5	88	98.5
Vicki	46	86	11.5	43	54.5	11.5	86	97.5
Standard deviation	4.0	2.0	1.0	1.0		1.0	2.0	
Relative weights			1	1		1	2	

given that the standard deviation of the scores in Set 2a is 2.24, estimate the standard deviation of scores in Sets 2b, 2c, and 2d. What are your answers?

	Scores	Standard deviation
Set 1a:	38 42 46	4.0
Set 1b:	138 142 146	?
Set 1c:	18 22 26	?
Set 1d:	18 20 22	?
Set 2a:	4 6 7 8 10	2.24
Set 2b:	5 7 8 9 11	?
Set 2c:	5 9 11 13 17	?
Set 2d:	10 18 22 26 34	?

Answers:

For Set 1b: 4.0. Variability between scores is the same as for Set 1a, even though scores are 100 points higher.

For Set 1c: 4.0. Variability between scores is the same as for Set 1b, even though scores are 120 points lower.

For Set 1d: 2.0. Variability between scores is half that of scores in Set 1c.

For Set 2b: 2.24. Variability between scores is the same as for Set 2a, even though scores are 1 point higher.

For Set 2c: 4.48. Variability between scores is twice that of scores in Set 2b.

For Set 2d: 8.96. Variability between scores is twice that of scores in Set 2c.

Note that the scores in Set 2d of Table 2 are simply the scores in Set 2c multiplied by two. When scores are multiplied by two, the variability of the scores is doubled, as is the standard deviation. However, when scores are changed by adding a constant (e.g., by adding 100 points to each score), the value of the standard deviation does not change. This brings us to a very important principle that can be used to control the weights of scores when assigning grades:

> When a set of scores is changed by adding (or subtracting) a constant, the variability of the scores does not change; correspondingly, standard deviation (the measure of variability) remains unchanged. But when a set of scores is changed by multiplying (or dividing) by a constant, the variability of scores does change proportionally, as does the value of the standard deviation. Therefore, multiplication and division can be used to give scores a desired standard deviation and likewise the intended weight.

Using Standard Deviation to Equalize Weights of Scores

If each set of scores has the same standard deviation, they will also have the same weight. This quality is applied to the scores in Table 2. As illustrated in Step 2, dividing scores on the first and second exams by 4 and 2 respectively results in their standard deviations' becoming one fourth and one half their original values. Now that the standard deviations of the two sets of scores are equal, the scores have equivalent weight. This is confirmed by the total scores. In a relative sense, all three students performed equally well on the two exams. Therefore, with the two scores weighted the same, these students obtained the same total score: 54.5.

In essence, the procedure used above consists of dividing each score by the standard deviation of the set of scores. Doing this always changes the standard deviation of scores to 1.0. When grades are being assigned by comparing students to each other, the weights of scores can be equated by dividing respective scores by their standard deviations.

Using Standard Deviations to Obtain Desired Weights of Scores

Once each set of scores has the same weight (i.e., equal standard deviations), the procedure for obtaining desired weights of these scores is very simple. One multiplies each set of scores by their desired weights. For example, if the second exam is to have twice the weight of the first exam, one multiplies each score on the first exam by one and each score on the second exam by two. This procedure is illustrated in Step 3 of Table 2.

Multiplying scores by desired weights is not an uncommon practice among instructors, but it is essential that scores be given equal weights before this multiplication occurs. Otherwise, the instructor is merely changing weights from one unknown value to another.

A Simpler Way to Control Standard Deviations

A convenient way to control the standard deviation of scores is to convert each score to a stanine. Stanines are scores that range from 1 to 9 and have a standard deviation of 2. If scores on each assignment and test are converted to stanines, each will have the same weight, because they have a common standard deviation. (This is illustrated later in Table 3.)

To convert scores to stanines, the scores for one assignment or test are ranked from high to low. A precise percentage of scores is then assigned to each stanine. Hills (1981) proposes the use of approximate percentages that are easy to remember, as each is a multiple of four. These percentages are given below (Example 5).

Stanine	Percentage
9	Highest 4%
8	Next highest 8%
7	Next highest 12%
6	Next highest 16%
5	Next highest 20%
4	Next highest 16%
3	Next highest 12%
2	Next highest 8%
1	Lowest 4%

If a class contained 100 students, the 4 highest scores would be assigned stanine 9, the 8 next highest scores stanine 8, the next 12 scores stanine 7, and so on. This same process would be repeated for scores on each assignment and test. For classes of other than 100 students, different numbers of scores are assigned each stanine.

The percentages given above in Example 5 often cannot be used when students receive tied scores or when classes do not contain a multiple of 25 students. However, as long as the percentage of scores assigned each stanine remains close to these values, the standard deviation of converted scores will be reasonably close to 2. After scores are given equal weights through conversion to stanines, they are multiplied by their desired weights.

Skill 3: Using the Appropriate Procedure for Weighting Scores When Assigning Course Grades

It is important to select an appropriate procedure for weighting scores, because alternate approaches can result in different grades' being assigned to students. This is illustrated by Parts A and B of Table 3. In Part A, grades are based on the percentage of points each student has earned. In this example, 90%, 80%, 70%, and 60% represent the minimum scores associated with A, B, C, and D, respectively. The first exam is intended to have twice the weight of the second. Before multiplying scores by their intended weights, exam scores are converted to percentages to equalize their maximum points (Step 2). Scores are then multiplied by their intended weights, totaled, and converted to total percentages (Step 3). Based on these percentages, 5 As, 8 Bs, 8 Cs, 3 Ds, and an F were assigned.

However, if grades are to be based on each student's *relative* performance, then the distribution of grades is predetermined, with the highest grades reserved for the students who obtain the highest total scores. To facilitate comparison to the previous example, it is assumed that the distribution of grades observed in Part A represents this predetermined distribution. Because grades are now to be assigned through comparison of students to each other, the standard deviation of scores, not the maximum points, controls their weights. In Part B of Table 3, exam scores are converted to stanines to equate their standard deviations (Step 2). These stanines are then multiplied by the desired weights and totaled (Step 3).

Note that the total scores resulting from these two procedures rank students differently. This change in rankings resulted in a grade change for 4 of the 25 students. To identify which weighting procedure is appropriate when determining course grades, a distinction must be made between criterion-referenced and norm-referenced grading.

Criterion-Referenced and Norm-Referenced Grading

In education, a useful distinction has been made between criterion-referenced and norm-referenced measurement. Glaser and Nitko (1971) indicate that a measure is criterion-referenced if it is "directly interpretable in terms of specified performance standards" (p. 653). Indicating that a student can type 20 words a minute is criterion-referenced because it states specifically what the student is able to do. No reference to the achievement of other students is needed to give meaning to the student's performance. On the other hand, a measure is norm-referenced if it describes what an individual can do relative to the performance of other students. Indicating that a student can type faster than 70% of the other students is norm-referenced. As with typing speed, a student's performance in a variety of skills can be expressed as both a criterion-referenced and a norm-referenced score.

Nitko (1984) explains that in order for scores to be criterion-referenced they must be referenced to a well-defined domain of tasks or behaviors. Therefore, scores interpreted in the context of "whatever a test measures," or simply in terms of being above or below a cut-off score, are *not* criterion-referenced.

Likewise, if the scores from a criterion-referenced test are presented in the absence of a well-defined domain, the scores no longer can be given a criterion-referenced interpretation. For instance, a student might obtain a score of 90% on a criterion-referenced test concerned with solving linear equations involving one unknown. However, simply stating that the student scored above 90% on an algebra test does not permit a criterion-referenced interpretation of that score. Similarly, assigning an A to students who exceed 90% in algebra does not allow criterion-referenced interpretations. Most course titles (e.g., Chemistry, English, and American History) are likewise too general to permit criterion-referenced interpretations of course grades, even if these grades are based on criterion-referenced measures.

Grades can be given norm-referenced interpretations, even when based on criterion-referenced measures. Norm-referenced interpretations are commonly given to letter grades. A student receiving a course grade of B is expected to be less proficient in the particular content area than a student who received an A, but more proficient than a student receiving a C. Although formal norms for course grades are usually not provided, grades are interpreted in the context of a reasonably well-known distribution.

Letter grades lend themselves to norm-referenced interpretations, but not without some concerns. For example, the lowest grade (e.g., F) is often used to indicate failure to achieve sufficient competence. Such grades should be assigned to students who will benefit from retention rather than distributed to a fixed proportion of students. Also, instructors who assign an atypical proportion of high or low grades limit the ability of others to interpret course grades.

Norm-referenced grading involves comparing each student to other students. The norm group can, but need not, consist of students currently enrolled in the course. As with standardized tests, norms can represent predetermined levels of performance established from students external to the present class. The remaining sections identify appropriate procedures for obtaining intended weights when (a) a class represents its own norm group and (b) students external to the class serve as the norm group. However, because it is assumed that grades in both situations represent normative comparisons, the standard deviation of scores and not the maximum number of points is used to control the weights of scores incorporated into the grade. Interestingly, the maximum number of points associated with a score will be shown to *facilitate* control of the standard deviations when students are compared to previously determined levels of performance.

Obtaining Intended Weights When a Class Represents Its Own Norm Group

If the ability of students enrolled in a course is about the same each term, the class can serve as its own norm group. Then it is defensible to use the same distribution of grades each term, because the achievement of students is reasonably constant. In this context, the proportion of students to receive each letter grade is determined at the beginning of the course. The following steps are then followed to give each score its intended weight.

Step 1. For each student, scores are obtained on each assignment and test to be incorporated into the course grade.

Step 2. Scores are given equal weights by equating their standard deviations. This can be accomplished by converting each set of scores to stanines.

Step 3. Each student's score is multiplied by its desired weight. These weighted scores are then totaled for each student.

TABLE 3

A Comparison of Grades Assigned When Weights of Scores Are Equated

Part A: Assigning Grades Using Percentages

	Step 1		Step 2		Step 3				
	Exam scores		Percentages		Weighted				
	1st	2nd	1st	2nd	1st	2nd	Total scores	Total percentages	Grade
Fred	25	17	100	85	200	85	285	95.0	A
Jason	24	18	96	90	192	90	282	94.0	A
Barbara	23	19	92	95	184	95	279	93.0	A
Isaac	24	17	96	85	192	85	277	92.3	A
Paul	22	19	88	95	176	95	271	90.3	A
Peggy	21	20	84	100	168	100	268	89.3	B
Stephanie	23	15	92	75	184	75	259	86.3	B
Anthony	21	18	84	90	168	90	258	86.0	B
Carla	21	18	84	90	168	90	258	86.0	B
Brian	21	15	84	75	168	75	243	81.0	B
Nancy	22	13	88	65	176	65	241	80.3	B
Lori	22	13	88	65	176	65	241	80.3	B
Dana	20	16	80	80	160	80	240	80.0	B
Tina	23	11	92	55	184	55	239	79.7	C
Stuart	20	15	80	75	160	75	235	78.3	C
Judith	19	16	76	80	152	80	232	77.3	C
Carol	20	14	80	70	160	70	230	76.7	C
Lee	22	9	88	45	176	45	221	73.7	C
Allyson	20	12	80	60	160	60	220	73.3	C
Chris	18	15	72	75	144	75	219	73.0	C
Anne	19	12	76	60	152	60	212	70.7	C
Joyce	21	8	84	40	168	40	208	69.3	D
Jean	19	11	76	55	152	55	207	69.0	D
Jamie	18	10	72	50	144	50	194	64.7	D
Dave	17	7	68	35	136	35	171	57.0	F
Maximum obtainable	25	20	100	100	200	100	300		
Relative weights			1	1	2	1			

Part B: Assigning Grades Using Stanines

	Step 1		Step 2		Step 3			
	Exam scores		Stanines		Weighted			
	1st	2nd	1st	2nd	1st	2nd	Total scores	Grade
Fred	25	17	9	6	18	6	24	A
Jason	24	18	8	7	16	7	23	A
Barbara	23	19	7	8	14	8	22	A
Isaac	24	17	8	6	16	6	22	A
Paul	22	19	6	8	12	8	20	A
Peggy	21	20	5	9	10	9	19	B
Stephanie	23	15	7	5	14	5	19	B
Anthony	21	18	5	7	10	7	17	B
Carla	21	18	5	7	10	7	17	B
Brian	21	15	5	5	10	5	15	B
Nancy	22	13	6	4	12	4	16	B
Lori	22	13	6	4	12	4	16	B
Dana	20	16	4	6	8	6	14	C*
Tina	23	11	7	3	14	3	17	B*
Stuart	20	15	4	5	8	5	13	C
Judith	19	16	3	6	6	6	12	C
Carol	20	14	4	5	8	5	13	C
Lee	22	9	6	2	12	2	14	C
Allyson	20	12	4	4	8	4	12	C
Chris	18	15	2	5	4	5	9	D*
Anne	19	12	3	4	6	4	10	C
Joyce	21	8	5	2	10	2	12	C*
Jean	19	11	3	3	6	3	9	D
Jamie	18	10	2	3	4	3	7	D
Dave	17	7	1	1	2	1	3	F
Average	21.0	14.0	5	5	10	5		
Standard deviation	2.0	3.6	2	2	4	2		
Relative weights			1	1	2	1		

*Grade has changed.

Obtaining Intended Weights When Comparing Students to an External Group

Instructors often prefer using previously established levels of performance as the basis for assigning grades. This may encourage students to learn cooperatively, because they are not competing against each other for high grades. Also, a single class is often too small to represent a reasonable norm group.

As described previously, stanines might be used to equate the weights of scores. Doing this, however, is impractical when the norm group is external to the class. Scores from previous students would have to serve as the basis of converting present scores to stanines, much as norms from standardized tests serve as the basis of converting raw scores to derived scores. The content, administration, and scoring of classroom tests and assignments would have to remain constant for these norms to be relevant.

A more practical approach is to select a level of performance for each grade which is expected to result in a distribution of grades similar to those assigned by other instructors at the school. Experienced instructors can estimate these levels quite effectively. Newer instructors often benefit from assistance when setting grading standards. All instructors should review the distribution of grades assigned and revise grading standards if the distribution varies considerably from that typically observed in other classes involving similar students.

Because levels of performance and, subsequently, grades are based on comparisons among students, weights of scores are controlled by their standard deviations. When students are graded through comparison to predetermined levels of performance, however, the standard associated with each grade is usually expressed as percentage of total points. We have found that the maximum number of possible points, not standard deviation, controls the weights of scores when grades are based on percentages.

Fortunately, with appropriate care, the standard deviation of scores will be proportional to the maximum number of points a student can obtain on each score. Consequently, when students are graded through comparison to predetermined performance levels, the maximum points associated with each score *can* be used to control its weight, and provide the same ranking of students that would be obtained had standard deviations been used to control weights. This is demonstrated below (Example 6).

	Test scores		
	1st	2nd	3rd
	25	50	50
	24	48	49
	23	46	48
	22	44	47
	22	44	47
	21	42	46
	20	40	45
	19	38	44
Maximum obtainable	25	50	50
Standard deviation	2	4	2

The maximum points *and* standard deviation of scores on Test 2 are twice those on Test 1. This happened because the distributions of scores on the two tests are proportional; scores range from approximately 75% to 100% of the maximum on both exams, with scores distributed the same within this range on both tests. *The standard deviation of scores will be proportional to the maximum points if the distribution of scores on each assignment and test remain proportional. If maximum points are proportional to standard deviations, maximum points can be substituted for standard deviations as the means for controlling weights of scores.* The proportional relationship between maximum points and standard deviation does not hold for Test 3 in Example 6. This is because the distribution of Test 3 scores is not proportional to distributions on the other two tests. Scores on Test 3 are distributed between approximately 90% to 100% of the maximum points.

Any distribution of scores can be used. For example, if scores on each test and assignment are distributed proportionally between 60% to 90%, standard deviations of scores will be proportional to the maximum points. In fact, each set of scores need not be exactly proportional. For instance, it is acceptable for one set of scores to range from 60% to 90% and another from 60% to 100%. This discrepancy would likely affect the standard deviations and therefore the weights of scores somewhat, but not significantly. On the other hand, if scores on a test range from 60% to 95% and scores on an assignment range from 90% to 100%, the standard deviations and thus the weights of the test and assignment would be far from proportional to their maximum obtainable points. Using this information, which assignment in the illustration below (Example 7) has approximately half the weight of the test?

		Assignments		
	Test	1	2	3
Maximum obtainable	20	10	10	10
	20	10	10	10
	19	10	10	8
	17	10	10	7
	16	9	10	6
	15	8	10	5
	14	8	9	5
	12	7	9	4
	11	7	9	4
	11	6	9	2
	10	6	8	1

Answer: Assignment 1. The maximum points a student can obtain on each of the assignments is half that of the test. However, the distribution of scores on only the first assignment are nearly proportional to those of the test. Scores are distributed from 50% to 100% on the test, and 60% to 100% on the first assignment.

When similar items are used on each classroom test, the scores on these tests often remain nearly proportional. When this is the case, the standard deviation and therefore the weight of scores on each test are proportional to the maximum points a student can obtain on the test. Consequently, doubling the number of items on a test typically doubles its weight. The variability of scores on assignments is usually more difficult to anticipate. For subjectively scored assignments, Oosterhof (in press) proposes a scoring scheme such as the one illustrated below (Example 8).

Score	Meaning
10	The very best in the class
9	Very good work
8	Good work
7	Acceptable, meets class standards

Descriptive statements are associated with points in a manner that will likely give the desired variability of scores. In this illustration, scores will likely be distributed between 70% to 100%, and the maximum obtainable number of points is 10. By associating different point values with each description, different maximum points and distributions of scores can be achieved.

In review, if grades are assigned by comparing students to an external group, performance levels are determined for each grade which are expected to result in a distribution of grades similar to that assigned by other instructors. The following steps are then followed to give each score its intended weight:

Step 1. Scores are obtained on each assignment for each student.

Step 2. If the distribution of scores on each assignment and test are fairly proportional, scores are given equal weights by equating their maximum points. This can be accomplished by converting scores to percentages.

Step 3. Each student's score is multiplied by its desired weight. These weighted scores are then totaled for each student, and grades are assigned based on the percentage of total points that each student has obtained.

These three steps can be combined by initially setting the maximum points obtainable on each assignment and test proportional to their desired weights. Then scores are simply totaled and grades assigned based on percentage of total points each student has achieved. However, care must still be taken to maintain a proportional distribution of scores on each test and assignment.

Summary

The maximum number of points a student can obtain on a test or assignment determines the weight of its scores when grades are based on percentage of points each student has achieved. In contrast, the weights of scores are affected by their standard deviation when grades are assigned through comparison of students to each other. Assuming that grades are used in a normative context, the weights of scores should be controlled through the adjustment of their standard deviations.

When a class represents its own norm group, the standard deviations and therefore weights of scores can be equated by converting scores to stanines. Once scores achieve equal weights, they are each multiplied by their desired weights. However, converting scores to stanines is impractical when students external to the class represent the norm group. Instead, grades are assigned by comparing students to predetermined performance levels. Weights are controlled by setting the maximum obtainable points on each test and assignment proportional to their desired weights. In this situation, the maximum points associated with each score will be proportional to standard deviations of scores and thus their weights *as long as* the distribution of scores is proportional on each test and assignment.

Teaching Aids Are Available

A set of teaching aids, designed by Albert C. Oosterhof to complement his ITEMS module, "Obtaining Intended Weights When Combining Students' Scores," is available at cost from NCME. These teaching aids consist of additional student exercises and overhead masters. As long as they are available, they can be obtained by sending $1.00 to: **Teaching Aids, ITEMS Module #2, NCME, 1230 17th St., NW, Washington, DC 20036.**

Notes

[1]Note that the value being equated is the maximum score that can be obtained (i.e., a perfect score). Quite possibly, this maximum score will not equal the highest score achieved by a student.

[2]For simplicity, many illustrations are based on a limited number of students. For instance, Example 3 implies that a class contains only three students. When grades are assigned through comparison to other students, the comparison group would typically involve considerably more students. Most illustrations also involve scores from only two assignments or exams. Typically, grades are based on several assignments, tests, and other observations. The principles discussed here generalize to settings involving greater numbers of students and scores.

[3]When three or more scores are involved, the relative weights of scores are affected by both the standard deviations of scores and the correlations among scores. These correlations are generally assumed not to significantly alter the weights of scores and are not included in discussions of course grades.

References

Glaser, R., & Nitko, A. J. (1971). Measurement in learning and instruction. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed., pp. 625–670). Washington, DC: American Council on Education.

Hills, J. R. (1981). *Measurement and evaluation in the classroom* (2nd ed.). Columbus, OH: Merrill.

Nitko, A. J. (1984). Defining "criterion-referenced test." In R. A. Berk (Ed.), *A guide to criterion-referenced test construction* (pp. 8–28). Baltimore, MD: Johns Hopkins University Press.

Oosterhof, A. C. (in press). *Classroom applications of educational measurement.* Columbus, OH: Merrill.

Additional Readings

K. F. Geisinger has written a selection in the fifth edition of the *Encyclopedia of Educational Research* concerned with marking systems (H. E. Mitzel (Ed.), 1982, New York: Free Press, pp. 1139–1145). This discussion reviews research on grading with emphasis on the purpose of grades, alternate marking systems, and psychometric qualities of grades.

Several of the texts directed at classroom applications of educational measurement include discussions on grading. The previously referenced books by Hills and Oosterhof emphasize student qualities that should be included in grades. J. S. Terwilliger has prepared a useful small book entitled *Assigning*

Grades to Students (1971, Glenview, IL: Scott, Foresman).

C. T. Holmes and K. M. Matthews conducted an extensive review of research regarding the effects of nonpromotion on elementary and junior high school students (1984, *Review of Educational Research, 54*, 225–236). Their findings are of particular relevance to marking, in that course grades often serve as the basis for determining whether a student will be held back from promotion.

Self-Test

The following illustration (Example 9) contains information concerning students' scores on five assignments.

	Assignment				
	1	2	3	4	5
Maximum possible points	10	20	15	40	30
Average score	7	15	13	30	24
Standard deviation of scores	2	3	1	6	5

1. Assume that course grades are to be based on the percentage of total points each student has obtained on these assignments. What will happen if scores are simply totaled without adjustment? Rank order these assignments with respect to the weight each has in determining the percentage of total points obtained by each student.

2. Now assume that course grades are to be based on each student's relative standing within the class. Again using information in Example 9, rank order the assignments in terms of the weight that each will have in determining students' relative standing.

3. If the distributions of scores on various assignments and tests are proportional, the standard deviations of these scores will be proportional to the maximum points students can obtain on each score. The following illustration (Example 10) lists scores obtained by 10 students on a test and five assignments. Which of the assignments have scores that are proportional to scores on the test?

	Test	Scores on assignments				
	scores	1	2	3	4	5
Maximum possible	50	10	20	40	50	50
	50	10	20	39	49	49
	49	10	20	36	45	49
	47	9	20	34	43	48
	46	9	20	33	42	47
	45	9	20	31	37	46
	45	9	20	30	35	45
	44	9	20	28	32	44
	43	8	20	26	29	42
	41	8	20	23	26	41
	40	8	16	20	20	40

Answers to the Self-Test

Items 1 and 2. The average score does not affect the relative weights of the scores. When grades are based on the percentage of total points, the maximum points a student can earn determines the relative weights of scores. Ranked from most to least weight, the assignments in Example 9 would be ordered 4, 5, 2, 3, and 1. When grades are based on students' relative position in class, the standard deviation of scores determines the relative weights of scores. Ranked from most to least weight, the assignments in Example 9 would now be ranked 4, 5, 2, 1, and 3.

Item 3. Scores on the test are distributed from 40 to 50, 80% to 100% of the maximum possible 50 points on the test. In order for scores on an assignment to be distributed proportionally, they will also have to range in the same manner from approximately 80% to 100%. Scores on Assignments 1 and 5 accomplish this. Scores on Assignment 2 range from 80% to 100% but are not proportionally distributed, in that all except the lowest score equal 100%. Scores on Assignment 3 are distributed from 50% to 100%, and scores on Assignment 4 are distributed from 40% to approximately 100%.

About ITEMS

The purpose of the Instructional Topics in Educational Measurement Series (ITEMS) is to improve the understanding of educational measurement principles. These materials are designed for use by college faculty and students as well as by workshop leaders and participants.

This series is the outcome of an NCME Task Force established in 1985 in response to a perceived need for materials to improve the communication and understanding of educational measurement principles. The committee is chaired by Al Oosterhof, Florida State University. Other members of the committee are Fred Brown, Iowa State University; Jason Millman, Cornell University; and Barbara S. Plake, University of Nebraska.

Topics for the series were identified from the results of a survey of a random sample of NCME members. Authors were selected from persons either responding to a call for authors that appeared in *Educational Measurement: Issues and Practice* or through individual contacts by the committee members. Currently, 17 authors are involved in developing modules. *EM* was selected as the dissemination vehicle for the ITEMS modules. Modules will appear, in a serial fashion, in future issues of *EM*. Barbara S. Plake is serving as editor of the series.

Each instructional unit consists of two parts, (1) instructional module and (2) teaching aids. The instructional modules, which will appear in *EM*, are designed to be learner-oriented. Each module consists of an abstract, tutorial content, a set of exercises including a self-test, and annotated references. The instructional modules are designed to be homogeneous in structure and length. The teaching aids, available at cost from NCME, are designed to complement the instructional modules in teaching and/or workshop settings. These aids will consist of tips for teaching, figures or masters from which instructors can produce transparencies, group demonstrations, additional annotated references, and/ or test items supplementing those included within the learner's instructional unit. The instructional module and teaching aids for an instructional unit are developed by the same author.

To maximize the availability and usefulness of the ITEMS materials, permission is hereby granted to make multiple photocopies of ITEMS materials for instructional purposes. The publication format of ITEMS in *EM* was specifically chosen with ease of photocopying in mind, as the modules appear in consecutive, text-dedicated pages.

ITEMS Module Teaching Aid

OBTAINING INTENDED WEIGHTS WHEN COMBINING STUDENTS' SCORES

Albert C. Oosterhof
Florida State University

The enclosed sheets can be photocopied for distribution and/or used to make overhead transparencies.

Compare the Standard Deviations

It is useful for participants to understand that standard deviation is a measures of score dispersion. This sheet lists pairs of score distributions and asks whether the standard deviation of the first distribution is the same, twice, or four times the standard deviation of the second distribution. Several examples illustrate that the magnitude of standard deviation and that of the average score can be independent.

Identify the Proportional Distributions

If students' scores on all assignments and tests are proportionally distributed, the standard deviations of scores will be proportional to the maximum points obtainable on each assignment and test. Consequently, the maximum points can be used to control the relative weights of the scores. Also, the weights of scores on different assignments and tests can then be equated by converting them to percents. Therefore it becomes important for participants to recognize when distributions of scores are approximately proportional.

Asking whether the following two conditions are present represents one technique for determining if distributions are proportional: 1) When converted to percents, do both distributions have approximately the same range (e.g., 60% to 100%)? 2) Are scores distributed about the same within that range (e.g. concentrated at the upper end of both distributions, or distributed quite evenly throughout the range of both distributions)?

Using the Class as Its Own Norm Group When Assigning Course Grades

Two sheets are associated with this exercise. The first list three scores for 25 students, and provides a work space for weighting and combining scores. The second sheet displays the completed exercise. Stanines are used here to equate the weights of scores. The ITEMS module describes how to convert scores to stanines.

Comparing Students to an External Group When Assigning Course Grades

Again, two sheets are associated with this exercise. The distribution of the 14 scores on the three assignments are approximately proportional. Therefore, the maximum points associated with each assignment is used to control their relative weights. Percents are used to equate the weights of scores, and then scores are multiplied by their desired weights.

Compare the Standard Deviations

Group 1 -	Test 1:	4 6 8 9 10 10 11 12 14 16
	Test 2:	3 5 7 8 9 9 10 11 13 15
Group 2 -	Test 1:	24 26 27 28 28 29 30 32
	Test 2:	14 16 17 18 18 19 20 22
Group 3 -	Test 1:	3 7 9 11 11 13 15 19
	Test 2:	7 9 10 11 11 12 13 15
Group 4 -	Test 1:	3 7 9 11 11 13 15 19
	Test 2:	17 19 20 21 21 22 23 25
Group 5 -	Test 1:	7 11 15 15 15 19 23
	Test 2:	13 14 15 15 15 16 17
Group 6 -	Test 1:	7 9 11 11 11 13 15
	Test 2:	13 14 15 15 15 16 17
Group 7 -	Test 1:	7 8 9 9 9 10 11
	Test 2:	13 14 15 15 15 16 17
Group 8 -	Test 1:	21 23 25 27 29
	Test 2:	10 12 14 16 18

Listed above are the scores different groups of students obtained on two tests. For each group of students, indicate whether the standard deviation of scores on the first test is the same, twice, or four times the standard deviation of scores on the second.

Answers: 1) same; 2) same; 3) twice; 4) twice; 5) four times; 6) twice; 7) same; 8) same.

Identify the Proportional Distributions

Example 1
A: (10) 7 8 9 9 10
B: (10) 7 8 8 9 10

Example 2
A: (10) 7 8 9 9 10
B: (20) 14 16 17 18 20

Example 3
A: (10) 7 8 9 9 10
B: (20) 7 11 14 15 20

Example 4
A: (10) 7 8 9 9 10
B: (10) 7 10 10 10 10

Example 5
A: (10) 5 7 8 9 10
B: (20) 11 14 15 17 20

Example 6
A: (10) 5 7 8 9 10
B: (20) 10 12 14 15 16

Example 7
A: (10) 5 7 8 9 10
B: (20) 15 17 18 19 20

Example 8
A: (10) 5 7 8 9 10
B: (20) 10 14 16 18 20

Example 9
A: (10) 5 7 8 9 10
B: (20) 10 15 16 19 20

Example 10
A: (20) 8 12 16 16 18
B: (30) 12 18 24 24 27

Example 11
A: (20) 8 12 16 16 18
B: (30) 12 19 23 24 28

Example 12
A: (20) 8 12 16 16 18
B: (30) 8 12 16 16 18

Example 13
A: (20) 15 17 17 18 19
B: (20) 15 15 20 20 20

Example 14
A: (20) 15 17 17 18 19
B: (20) 15 17 18 18 20

If students' scores on all assignments and tests are proportionally distributed, the standard deviations of scores will be proportional to the maximum points. Consequently, the maximum points can be used to control the relative weights of the scores.

Listed above are the scores of five students on pairs of assignments (A and B). For each pair, indicate whether the scores on the two assignments are <u>approximately</u> proportional. The maximum possible score a student could obtain on each assignment is given within parentheses. For instance, the maximum possible score on assignment B within Example 6 is 20 points. The highest score a student obtained on that assignment was 16 points, 80% of the maximum.

Answers: 1) yes; 2) yes; 3) no; 4) no; 5) yes; 6) no; 7) no; 8) yes; 9) yes; 10) yes; 11) yes; 12) no; 13) no; 14) yes.

Using the Class as Its Own Norm Group When Assigning Course Grades

	(Step 1) Assignment Scores			(Step 2) Stanines			(Step 3) Weighted				
	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd	Total	Grade
Michael	25	12	29	__	__	__	__	__	__	__	__
Leiana	24	13	32	__	__	__	__	__	__	__	__
Daniel	24	11	31	__	__	__	__	__	__	__	__
Tom	23	12	35	__	__	__	__	__	__	__	__
Ann	23	14	31	__	__	__	__	__	__	__	__
Frances	22	10	26	__	__	__	__	__	__	__	__
Victoria	21	12	32	__	__	__	__	__	__	__	__
Heidi	19	13	28	__	__	__	__	__	__	__	__
Barry	19	11	26	__	__	__	__	__	__	__	__
Pamela	19	10	24	__	__	__	__	__	__	__	__
Richard	18	10	30	__	__	__	__	__	__	__	__
Scott	18	11	25	__	__	__	__	__	__	__	__
James	17	9	22	__	__	__	__	__	__	__	__
Laura	17	10	27	__	__	__	__	__	__	__	__
Camille	17	8	25	__	__	__	__	__	__	__	__
Joan	16	11	24	__	__	__	__	__	__	__	__
Gail	15	9	22	__	__	__	__	__	__	__	__
Jose	15	10	19	__	__	__	__	__	__	__	__
Ada	15	9	27	__	__	__	__	__	__	__	__
Terry	14	8	23	__	__	__	__	__	__	__	__
Neal	14	7	15	__	__	__	__	__	__	__	__
Steven	14	8	25	__	__	__	__	__	__	__	__
Tammy	13	6	23	__	__	__	__	__	__	__	__
Rita	11	9	21	__	__	__	__	__	__	__	__
Paula	8	7	18	__	__	__	__	__	__	__	__
Desired weights				1	1	1	2	3	5		

Step 1: Students' scores on three assignments are provided.

Step 2: Give the scores equal weights by converting them to stanines. (The standard deviation of scores on each assignment becomes equal, therefore the scores obtain equal weight.)

Step 3: Now that the scores have equal weights, multiply the respective scores by their desired weights. (In this example, the desired weights are 2, 3, and 5.) Total the scores and assign 4 A's, 10 B's, 8 C's, and 3 D's.

Using the Class as Its Own Norm Group When Assigning Course Grades

	(Step 1)			(Step 2)			(Step 3)				
	Assignment Scores			Stanines			Weighted				
	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd	Total	Grade
Michael	25	12	29	9	7	6	18	21	30	69	B
Leiana	24	13	32	8	8	8	16	24	40	80	A
Daniel	24	11	31	8	6	7	16	18	35	69	B
Tom	23	12	35	7	7	9	14	21	45	80	A
Ann	23	14	31	7	9	7	14	27	35	76	A
Frances	22	10	26	7	5	5	14	15	25	54	B
Victoria	21	12	32	6	7	8	12	21	40	73	A
Heidi	19	13	28	6	8	6	12	24	30	66	B
Barry	19	11	26	6	6	5	12	18	25	55	B
Pamela	19	10	24	6	5	4	12	15	20	47	B
Richard	18	10	30	5	5	7	10	15	35	60	B
Scott	18	11	25	5	6	5	10	18	25	53	B
James	17	9	22	5	4	3	10	12	15	37	C
Laura	17	10	27	5	5	6	10	15	30	55	B
Camille	17	8	25	5	3	5	10	9	25	44	C
Joan	16	11	24	4	6	4	8	18	20	46	C
Gail	15	9	22	4	4	3	8	12	15	35	C
Jose	15	10	19	4	5	2	8	15	10	33	C
Ada	15	9	27	4	4	6	8	12	30	50	B
Terry	14	8	23	3	3	4	6	9	20	35	C
Neal	14	7	15	3	2	1	6	6	5	17	D
Steven	14	8	25	3	3	5	6	9	25	40	C
Tammy	13	6	23	2	1	4	4	3	20	27	D
Rita	11	9	21	2	4	3	4	12	15	31	C
Paula	8	7	18	1	2	2	2	6	10	18	D
Mean	17.6	10.0	25.6	5	5	5	10	15	25		
Standard Dev.	4.3	2.0	4.8	2	2	2	4	6	10		
Relative Weight				1	1	1	2	3	5		

Comparing Students to an External Group When Assigning Course Grades

	(1st Step)			(2nd Step)			(3rd Step)					
	Scores on Assignments			Percent Scores			Weighted Scores					
	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd	Total	%	Grade
Nelson	10	22	17									
Tara	10	21	20									
Carla	9	24	20									
Anthony	9	22	16									
Cheryl	9	23	14									
Leslie	8	22	18									
Gregg	8	20	16									
Linda	8	22	14									
Chad	8	20	18									
Teresa	8	19	15									
Valerie	7	22	14									
Russell	7	19	15									
Robin	7	14	13									
Adam	6	15	14									
Maximums...	10	25	20									
Desired Weights...				1	1	1	3	2	5			

Step 1: Students' scores on three assignments are provided, as are the maximum points a student could obtain on these assignments.

Step 2: Give the scores equal weights by converting them to percents. (The maximum possible score on assignments becomes equal, therefore their scores obtain equal weight.)

Step 3: Now that the scores have equal weights, multiply the respective scores by their desired weights. (In this example, the desired weights are 3, 2, and 5.) Total the scores and convert the totals to percents. Assign grades using 90-100% = A, 80-90% = B, 70-80% = C, and 60-70% = D.

Comparing Students to an External Group When Assigning Course Grades

	(1st Step) Scores on Assignments			(2nd Step) Percent Scores			(3rd Step) Weighted Scores					
	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd	Total	%	Grade
Nelson	10	22	17	100	88	85	300	176	425	901	90%	A
Tara	10	21	20	100	84	100	300	168	500	968	97%	A
Carla	9	24	20	90	96	100	270	192	500	962	96%	A
Anthony	9	22	16	90	88	80	270	176	400	846	85%	B
Cheryl	9	23	14	90	92	70	270	184	350	804	80%	B
Leslie	8	22	18	80	88	90	240	176	450	866	87%	B
Gregg	8	20	16	80	80	80	240	160	400	800	80%	B
Linda	8	22	14	80	88	70	240	176	350	766	77%	C
Chad	8	20	18	80	80	90	240	160	450	850	85%	B
Teresa	8	19	15	80	76	75	240	152	375	767	77%	C
Valerie	7	22	14	70	88	70	210	176	350	736	74%	C
Russell	7	19	15	70	76	75	210	152	375	737	74%	C
Robin	7	14	13	70	56	65	210	112	325	647	65%	D
Adam	6	15	14	60	60	70	180	120	350	650	65%	D
Maximums...	10	25	20	100	100	100	300	200	500	1000		
Relative Weights...				1	1	1	3	2	5			