Impact of End-of-Course Tests on Accountability Decisions in Mathematics and Science

A paper presented at the annual meeting of the National Council on Measurement in Education
New Orleans, LA

Paula Cunningham

Anthony Fina

William Adams

Catherine Welch

April 2011

Abstract

Students from four urban school districts were administered end-of-course tests in Algebra I and in Biology as well as the general assessment used by the state for accountability decisions. The end-of-course test scores were combined with scores on the corresponding subtests of the general assessment by forming composite scores in which the two tests were weighted differentially. The proficiency of the students in the study was determined based on composite scores to show how the addition of scores on end-of-course Algebra I and Biology tests can affect proficiency. End-of-course tests may prove a valuable addition for enhancing the information available for accountability decisions beyond what is offered by the general assessment alone.

Background and Purpose

States use assessments to support student learning, make accountability decisions, and assess readiness for post-secondary education and training (Vranek, 2008). Two types of assessments are commonly used: comprehensive tests that assess a broad range of content in particular subject areas and end-of-course tests that assess mastery of the standards for specific high school courses after their completion (Center on Education Policy, 2008; Vranek, 2008). The growing popularity of end-of-course tests reflects some discontent with the use of comprehensive exams in high school, where end-of-course tests are perceived as being more useful than comprehensive tests in helping educators modify curriculum and instruction (Gewertz, 2007). In addition to being more sensitive to instruction, end-of-course tests can tap higher level content and skills, may align better to the high school curriculum, and can help ensure consistency of rigor in courses taught throughout a state (Achieve, Inc., 2007, 2008). Testing immediately after completion of the course allows educators to teach the subject more deeply, and end-of-course tests may offer better assessment of mastery of specific content and a clearer picture of what was learned during the course (Gewertz, 2007; Center on Education Policy, 2008).

Increased interest in the use of end-of-course tests has led to a consideration of their place within a balanced system of assessment. The end-of-course test may provide further information about student achievement than that provided by the comprehensive assessment alone. In order to have a positive impact on learners, assessments should move beyond simple judgments about proficiency to provide richer descriptions of student performance (Stiggins, 2006). The use of multiple measures of a construct is essential to achieving a more balanced system of assessment. Whereas the comprehensive assessment influences decisions mostly at the program and policy level, the end-of-course assessment has its impact at the instructional support level and at the classroom level (Center on Education Policy, 2008; Stiggins, 2006).

Comprehensive assessments are still far more common than end-of-course tests, with only 16 states using end-of-course tests in their systems of assessment and 11 more planning to incorporate them in the future. Although some states that have end-of-course tests do not use them for the purpose of

accountability, 12 states use or plan to use at least some of their end-of-course tests to meet accountability requirements (Vranek, 2008). Some states are moving completely to end-of-course tests, while others will blend end-of-course exams with comprehensive tests (NGA Center for Best Practices, 2008).

In addition to the current accountability model, the reauthorization of the ESEA blueprint will likely require the implementation of college- and career-ready standards and the development of assessments aligned with those standards. The high school portion of the college and career standards specify the mathematics that all students should know in areas such as algebra, geometry, and probability and statistics. These standards align directly with the curriculum found in traditional high school courses and may be best assessed through end-of-course tests. This investigation is a pilot study to examine how scores on end-of-course assessments can add information about proficiency beyond that provided by scores on a comprehensive test.

In this study, four districts in a consortium of the eight largest school districts in a Midwestern state administered end-of-course assessments in Algebra I and Biology in the spring of 2010. One goal of the study is to describe how many of the student participants were considered proficient in mathematics and science on the most recent general assessment and how many of them would be considered proficient in mathematics and science if the end-of-course tests in Algebra I and Biology were used for accountability instead. A further objective of the study is to consider how the end-of-course scores in Algebra I and in Biology can be combined with the mathematics and science scores, respectively, on the comprehensive general assessment currently in use for accountability, as well as the reliability of the resulting composite scores. Finally, the proficiency of students in the study was evaluated based on composite scores obtained by different methods to show how the addition of scores on end-of-course Algebra I and Biology tests can affect proficiency.

Sources of Information

Subjects

Students were drawn from four urban school districts in a Midwestern state. Accordingly, it was not expected that the sample would be reflective of the statewide distribution of students. The study sample includes 1,188 students, 553 of whom took the end-of-course assessment in Algebra I and 635 of whom took the end-of-course assessment in Biology. See Table 1 for a demographic breakdown of the sample.

Chi-square goodness of fit tests conducted between the population of participants and the statewide student population showed that the pilot study subjects differed significantly from the state population on sex ($p=.011$), percentage of African-American students ($p<.001$), percentage of Hispanic students ($p=.003$), percentage of white students ($p<.001$), socio-economic status (SES, $p<.001$), English language learner status (ELL, $p<.001$), and individualized education program (IEP, $p<.001$) variables.

The study group was 47.6% male, 52.4% female, 14.3% African-American, 7.9% Hispanic, and 75.3% white; 42.0% qualified for free or reduced lunch programs, 5.2% were English language learners, and 9.0% had an IEP. The statewide student population is 51.2% male, 48.7% female, 5.2% African-American, 5.9% Hispanic, and 87.4% white; 30.8% qualify for free or reduced lunch programs, 2.8% are English language learners, and 13.0% have an IEP (Iowa Department of Education, 2009a).

Table 1

Demographic Data of Study Participants

Breakdown of Pilot Study Sample	N=1,188
Sex	
Male	565
Female	623
Race	
African-American	170
American Indian	9
Asian	20
Hispanic	94
White	894
Other Demographic Variables	
ELL	62
Migrant	8
SES Eligible	499
IEP	107

Note. ELL=English language learner. SES Eligible refers to qualifying for free or reduced price student lunches. IEP=Individualized educational program.

Instruments

The end-of-course (EOC) assessments in Algebra I and in Biology measure how well students have met the academic standards of those particular high school courses. Course grades in either Algebra I or Biology were also obtained for the participants in this study. The general assessment used by the state for accountability decisions is a standardized achievement battery.

Based on the sample of study participants, summary statistics were computed for each EOC test (KR-20, mean percent correct score, standard deviation, and Pearson correlation of EOC raw scores with the corresponding general assessment subtest raw scores). These statistics are reported in Table 2. It is evident from the mean percent correct scores that the students found the EOC tests challenging, and in

particular student performance was low on the test covering Algebra I. The difficulty of the Algebra I test is perhaps surprising (mean percent correct score=40), yet this result is consistent with student performance on the Algebra II test developed by the American Diploma Project (ADP), on which mean scores ranged from 21 percent correct to 35 percent correct across states (Achieve, Inc, 2008).

Table 2

Summary statistics for the EOC assessments in Algebra I and Biology

EOC Test	N	KR-20	$\overline{X}$	SD	r
Algebra I	553	.639	40.24	13.55	.501
Biology	635	.814	59.60	18.51	.665

Note. $\overline{X}$ is the mean percent correct score; r is the Pearson correlation between the EOC raw score and the subtest raw score on the assessment used by the state for accountability. Values of KR-20 obtained for samples much larger than that of the present study were .793 for Algebra I and .795 for Biology.

By way of comparison, the mean percent correct scores on the general assessment subtests in mathematics and science were 52.92 (SD=14.42) and 60.92 (SD=17.56), respectively.

The modest correlations reported in Table 2 (.501 and .665 for Algebra I and Biology, respectively) demonstrate that the EOC tests and the general assessment subtests measure slightly different constructs, which suggests that the EOC tests can add information about student achievement and the determination of proficiency if they are incorporated in some manner into a composite score with the general assessment.

Table 3

Correlations among the Mathematics Subtest of the General Assessment, the EOC Algebra I Test, and Algebra I Course Grades

	Mathematics	EOC Algebra I	Course Grade
Mathematics	1.000		
EOC Algebra I	.501	1.000	
Course Grade	.307	.394	1.000

Pearson correlation values for course grades in Algebra I with the mathematics subtest of the general assessment and the EOC Algebra I test are included in Table 3. Correlations for the course grades in Biology with the science subtest of the general assessment and the EOC Biology test are shown in Table 4. The low correlations found between course grades and the EOC test scores may result from the fact that grading standards differed for every teacher. Some teachers included the EOC test score as part of the course grade, while others did not. Some teachers may have factored in projects and presentations into the grade to a greater or lesser degree, while others may have used only homework and examination scores to determine course grades. The low correlations between each course grade and the corresponding subtest of the general assessment are the result of the general assessment covering a wider range of content than that covered by the single course of Algebra I or Biology.

Table 4

Correlations among the Science Subtest of the General Assessment, the EOC Biology Test, and Biology Course Grades

	Science	EOC Biology	Course Grade
Science	1.000		
EOC Biology	.665	1.000	
Course Grade	.492	.579	1.000

Methods

Composite Scores

In dealing with combinations of test scores rather than single test scores, a composite score may be viewed as the algebraic sum of weighted scores. The composite score derived from two tests for a person p can be denoted Z_p, where $Z_p = w_1X_{p1} + w_2X_{p2}$ (Haertel, 2006).

Several compensatory composite scores (Kane & Case, 2004) for combining the EOC Algebra I and Biology tests with the corresponding subtests on the state's general assessment were developed (Table 5). First is a composite score where the scores for the EOC test and the general assessment subtest are weighted equally. The second composite score is defined by simply adding the items from the two

tests. A third composite score is based upon the amount of time it takes to administer each of the two tests. The final composite score is determined by the distribution of the combined tests' items over the content strands of the state curriculum in ninth grade mathematics and tenth grade science (Iowa Department of Education, 2009b, 2010).

Table 5

Derivation of Composite Scores

Weighting	Composite Score, *Z*
Mathematics/Algebra I	
Equal	$.5X_{General}+.5X_{EOC}$
Number of items	$(81/111)X_{General}+(30/111)X_{EOC}$
Testing time	$(60/100)X_{General}+(40/100)X_{EOC}$
Content	$(53/111)X_{Strand1}+(18/111)X_{Strand2}+(10/111)X_{Strand3}+(30/111)X_{Strand4}$
Science/Biology	
Equal	$.5X_{General}+.5X_{EOC}$
Number of items	$(43/73)X_{General}+(30/73)X_{EOC}$
Testing time	$(30/70)X_{General}+(40/70)X_{EOC}$
Content	$(20/73)X_{Strand1}+(38/73)X_{Strand2}+(7/73)X_{Strand3}+(8/73)X_{Strand4}$

Reliability of composites. Estimates of the reliability of composites formed from congeneric parts of unequal lengths can be obtained from Raju's coefficient,

$$\rho_{XX'}=\frac{\sigma_X^2-\Sigma\sigma_{X_i}^2}{(1-\Sigma\lambda_i^2)\sigma_X^2}$$

where λ_i is the proportion of total test length for part-test i, and $\Sigma\lambda_i=1$ (Haertel, 2006). The first three composites described above derive from two unequal parts, with λ_1 and λ_2 representing the proportions attributable to the state accountability test and to the EOC test, respectively. The final composite is comprised of four parts, consisting of items on both tests that fall within the four content strands of the state standards. Table 6 shows the distribution of items on the state assessment and the EOC tests over the content strands of the Iowa Core Curriculum.

Table 6

Distribution of Items on the State Accountability Subtests and on the EOC Tests in Algebra I and Biology over the Content Strands of the Iowa Core Curriculum

Content Strands	# Items	Weights
Mathematics:		
Algebra	53	$\lambda_1 = 53/111$
Data Analysis/Stat. & Prob.	18	$\lambda_2 = 18/111$
Geometry & Measurement	10	$\lambda_3 = 10/111$
Quantitative Literacy	30	$\lambda_4 = 30/111$
Science:		
Science as Inquiry	20	$\lambda_1 = 20/73$
Life Science	38	$\lambda_2 = 38/73$
Earth & Space	7	$\lambda_3 = 7/73$
Physical Science	8	$\lambda_4 = 8/73$

Cut Points for Proficiency

Student proficiency on the general assessment was established by a standard setting that determined raw score cut points on the subtests. These cut points were determined prior to this study and are currently used for determining proficiency.

Proficiency on the EOC tests was considered in several ways. First, a cut point on each EOC test was set at a percent correct score corresponding to the same point on the subtests of mathematics and science on the state accountability assessment (raw score=12). Second, a raw score cut point was considered at the fifty-percent correct score for each EOC test (raw score=15). Third, a panel of content experts in the appropriate field conducted a preliminary standard setting according to the bookmark procedure and set proficiency cut scores for the EOC tests in three rounds (raw scores=18, 20, 22). In both Algebra I and Biology, the effects of different composites on student proficiency were evaluated using the raw score cut points for the general assessment subtests and the cut points for each EOC test. The tests were weighted in the composites by equal parts, by testing time, and by weights determined to maximize the reliability of the composite.

Maximizing the reliability of a composite. A general formula for the reliability of a composite of two parts as a function of the weights, component reliabilities, and the correlation between the components is

$$\rho_{LL'} = \frac{w_1^2\rho_{11'} + w_2^2\rho_{22'} + 2w_1w_2\rho_{12}}{w_1^2 + w_2^2 + 2w_1w_2\rho_{12}}$$

where the weights for maximizing the composite score reliability can be obtained by setting the first derivative with respect to w_1/w_2 to zero and solving (Rudner, 2001) or by graphing the function with respect to all values of w_1.

Results

Proficiency of Study Participants

Of the students taking the EOC Algebra I test, 67.8% were proficient in mathematics on the most recent general assessment. Of those taking the EOC Biology test, 78.6% were proficient in science on their most recent general assessment. Statewide 76.0% of students taking the general assessment were proficient in mathematics while 81.6% of students were proficient in science (Iowa Department of Education, 2009a). Proficiency of the participants on the EOC Algebra I and Biology tests was determined by (a) raw score cut points that correspond to the percent correct cut points on the general assessment subtests and (b) raw score cut points established by three rounds of standard setting using the bookmark procedure. The results of this analysis for Algebra I and Biology are presented in Tables 7 and 8, respectively. From these tables it is clear that if the EOC tests alone were used for accountability decisions, levels of proficiency would be dramatically different (considerably lower, for the most part) from those obtained using the general assessment.

Table 7

Percent of Students Proficient on the Mathematics Subtest of the General Assessment and on the EOC Algebra I Test at Several Raw Score Cut Points

Test	% Proficient
Mathematics	67.8
EOC Percent Correct	
RS=12	52.1
RS=15	27.3
EOC Expert Judgment	
RS=18	9.8
RS=20	4.9
RS=22	2.4

Note. RS=raw score. EOC Percent Correct refers to cut points on the EOC test corresponding to the percent correct cut points on the general assessment and to the fifty percent correct point; EOC Expert Judgment refers to the cut points established by three rounds of standard setting by the bookmark procedure.

Table 8

Percent of Students Proficient on the Science Subtest of the General Assessment and on the EOC Biology Test at Several Raw Score Cut Points

Test	% Proficient
Science	78.6
EOC Percent Correct	
RS=12	84.6
RS=15	71.0
EOC Expert Judgment	
RS=18	52.9
RS=20	43.0
RS=22	29.6

Note. RS=raw score. EOC Percent Correct refers to cut points on the EOC test corresponding to the percent correct cut points on the general assessment and to the fifty percent correct point; EOC Expert Judgment refers to the cut points established by three rounds of standard setting by the bookmark procedure.

Although most students were classified as proficient or not proficient according to both tests, some students were proficient on one test but not on the other. For the raw score cut point of 20 on the EOC Biology test, 232 out of 629 students (36.9%) were advantaged by the state assessment used for accountability, with only 8 (1.3%) disadvantaged by it. Figure 1 depicts the impact of the cut scores on proficiency, showing the distribution of test scores with the cut scores indicated by lines. Likewise, using the raw scale cut point of 12 on the EOC Biology test yielded 81 out of 629 (12.9%) advantaged by the EOC test with 44 (7.0%) disadvantaged by it. Figure 2 shows the result when the cut score on the EOC Biology test is shifted downward.

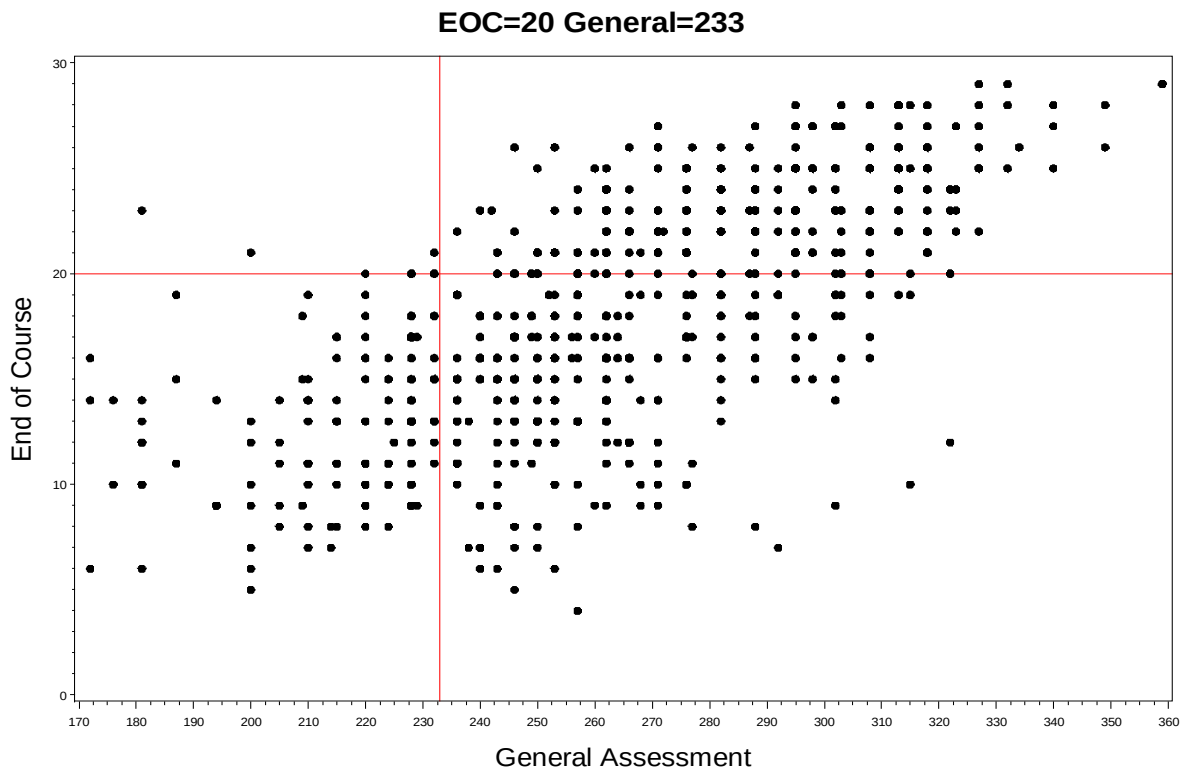

Figure 1. Impact on Proficiency of a Cut Score of 20 on the EOC Biology Test

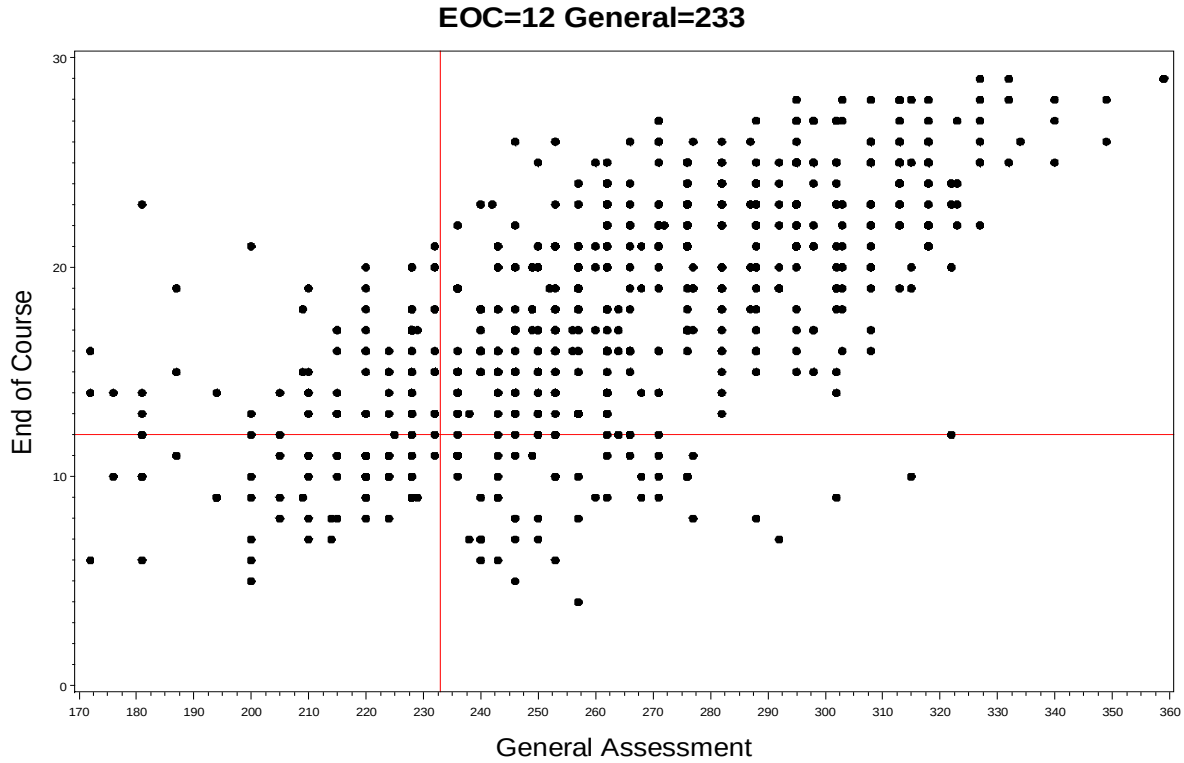

Figure 2. Impact on Proficiency of a Cut Score of 12 on the EOC Biology Test

Decision consistency. A decision consistency index φ, or proportion of consistent decisions, was calculated for the mathematics subtest of the general assessment at its proficiency cut point and the EOC Algebra I test at each of the five raw score cut points considered above. These results are presented in Table 9.

Table 9

Decision Consistency between the Mathematics Subtest of the General Assessment and the EOC Algebra I Test

Cut Score on EOC Algebra I Test	φ (Proportion of Consistent Decisions)
Percent Correct	
RS = 12	.678
RS = 15	.527
Expert Judgment	
RS = 18	.409
RS = 20	.373
RS = 22	.351

Note. RS=raw score. EOC Percent Correct refers to cut points on the EOC test corresponding to the percent correct cut points on the general assessment and to the fifty percent correct point; EOC Expert Judgment refers to the cut points established by three rounds of standard setting by the bookmark procedure.

The decision consistency index φ was also calculated for the science subtest of the general assessment at its proficiency cut point and the EOC Biology test at five raw score cut points (Table 10). Values of φ decreased as the raw score cut points for each EOC test increased, and the proportion of consistent decisions was uniformly higher for the EOC Biology test than for the Algebra I test, reflecting the fact that the EOC Algebra I test was quite difficult.

Table 10

Decision Consistency between the Science Subtest of the General Assessment and the EOC Biology Test

Cut Score on EOC Biology Test	φ (Proportion of Consistent Decisions)
Percent Correct	
RS = 12	.803
RS = 15	.799
Expert Judgment	
RS = 18	.725
RS = 20	.665
RS = 22	.551

Note. RS=raw score. EOC Percent Correct refers to cut points on the EOC test corresponding to the percent correct cut points on the general assessment and to the fifty percent correct point; EOC Expert Judgment refers to the cut points established by three rounds of standard setting by the bookmark procedure.

Composite Score Reliabilities

Reliabilities were calculated using Raju's coefficient (Haertel, 2006) for four types of scores: the composite obtained by weighting the general assessment subtest and the EOC test equally, the composite formed through weighting by the number of items in each test, the composite obtained when parts are weighted by the proportion of testing time required for each test, and the composite obtained by weighting parts representing content strands by the number of test items measuring those content areas.

The results of reliability analysis for composite scores formed using the mathematics subtest of the general assessment and the EOC Algebra I test are summarized in Table 11. The calculated reliability was noticeably higher for the case in which the two tests were weighted by the proportion of items

accounted for by each test, where λ_1=.73 and λ_2=.27. The reliabilities calculated using the other weighting schemes were very close to one another.

Table 11

Reliabilities of Composite Scores of the Mathematics Subtest with the EOC Algebra I Test

Weighted by	λ_1	λ_2	λ_3	λ_4	$\rho_{XX'}$
Equal	.5	.5			.639
Number of items	81/111	30/111			.810
Testing time	60/100	40/100			.665
Content	53/111	18/111	10/111	30/111	.683

Results were also obtained for composite scores formed using the science subtest of the general assessment and the EOC Biology test, and these are summarized in Table 12. Again the calculated reliability was highest when the two tests were weighted by the proportion of items within the composite, where λ_1=.59 and λ_2=.41. For these tests, the reliabilities calculated according to the other weighting schemes were close to one another, as was the case for the mathematics subtest and EOC Algebra I test.

Table 12

Reliabilities of Composite Scores of the Science Subtest with the EOC Biology Test

Weighted by	λ_1	λ_2	λ_3	λ_4	$\rho_{XX'}$
Equal	.5	.5			.776
Number of items	43/73	30/73			.801
Testing time	30/70	40/70			.792
Content	20/73	38/73	7/73	8/73	.786

Weights for maximizing the reliability of composite scores. The reliability of the observed composite score as proposed by Rudner (2001) is depicted graphically in Figure 3 for the EOC Algebra I and Biology tests with their corresponding subtests on the general assessment. The curves show the entire range of reliability for all possible weightings of the general assessment in the composite scores. Since the weights for the two parts sum to 1, as the weight of the general assessment in the composite increases, the weight for the EOC test decreases in kind (Kane & Case, 2004). For these calculations, the reliabilities of the general assessment subtests in mathematics and science were .929 and .889,

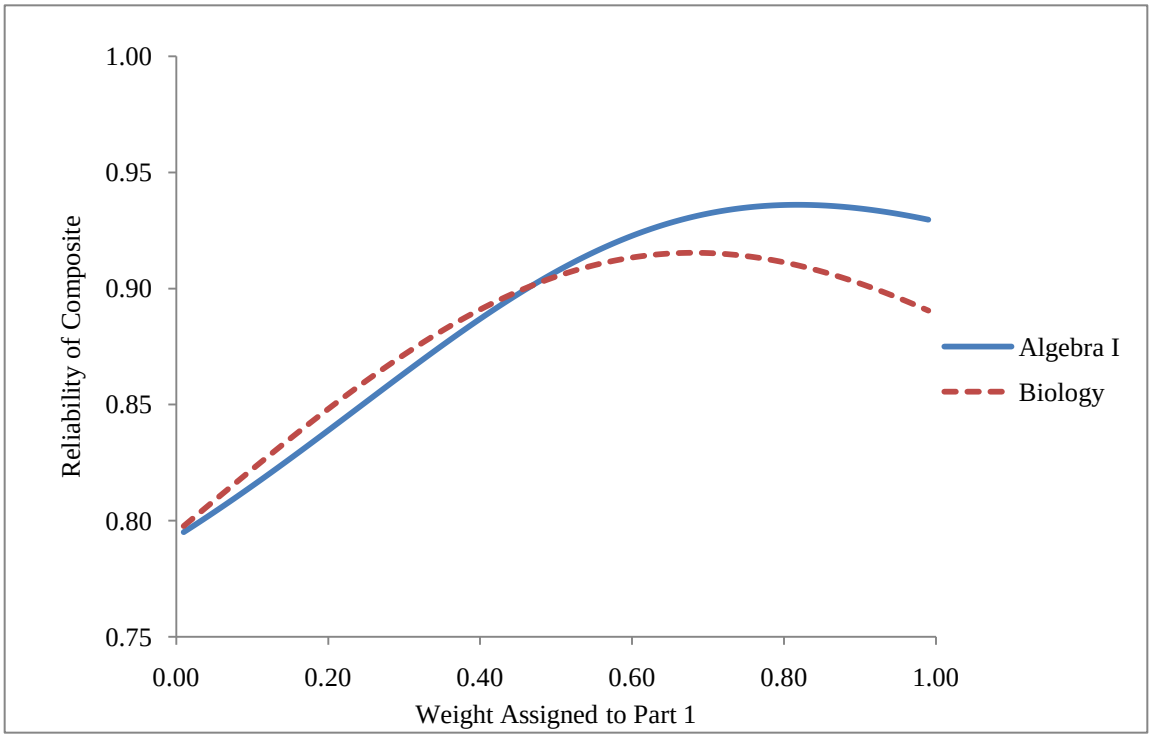

Figure 3. Reliability versus Weight of General Assessment (w_1)

respectively; those for the EOC tests in Algebra I and Biology were .793 and .795, respectively; and the correlations between tests for mathematics and science subject areas were .501 and .665, respectively. For the composite formed by the mathematics subtest and the EOC Algebra I test, w_1=.82 and w_2=.18 maximized reliability at .936; for the composite formed by the science subtest and the EOC Biology test, weights that maximized reliability at .915 were w_1=.68 and w_2=.32.

Proficiency Based on Composite Scores

In each subject area, the impact on proficiency of several types of composite scores was evaluated using the raw score cut points for the general assessment subtests and the cut points for each EOC test. Proficiency was determined for composites weighted by equal parts, by testing time, and by weights determined to maximize the composite's reliability.

Results for the composites formed from the weighting of the mathematics subtest of the general assessment and the EOC Algebra I test are shown in Table 13. As expected with the cut score on the general assessment subtest held constant, student proficiency decreased as the cut score on the EOC Algebra I test was increased. It was found that weighting by testing time (w_1=.6 and w_2=.4) gave higher proficiencies than weighting by equal parts (w_1=w_2=.5), and the highest proficiencies were obtained when the two tests were weighted to maximize reliability of the composite (w_1=.82 and w_2=.18). It is important to note that student proficiency on the mathematics subtest of the general assessment alone was 67.8%.

Table 13

Percent Proficient on Several Composite Scores at Various EOC Algebra I Test Cut Points

Cut Score on EOC Algebra I	Weighting by Equal Parts	Weighting by Testing Time	Weighting to Maximize Reliability
Percent Correct			
RS = 12	69.8	71.2	73.1
RS = 15	64.2	66.7	68.9
Expert Judgment			
RS = 18	53.2	61.5	68.9
RS = 20	47.7	57.0	65.1
RS = 22	43.6	50.8	65.1

Note. RS=raw score.

The results for student proficiency on the science composites at various EOC Biology test cut points are shown in Table 14. Again as expected, proficiency decreased as the cut score on the EOC Biology test increased with the general assessment subtest cut score held constant. There was no consistent variation of proficiencies between the different types of composite scores, however. Weighting by equal parts (w_1=w_2=.5) gave the highest proficiency at the lowest raw score cut point on the EOC Biology test. Weighting by testing time (w_1=.43 and w_2=.57) gave the highest proficiency only at the fifty-percent correct cut point. Finally, weighting to maximize reliability of the composite score (w_1=.68 and w_2=.32) gave the highest proficiencies of the various composites only at the high raw score cut points

determined by the preliminary standard setting. For reference, student proficiency on the science subtest of the general assessment alone was 78.6%.

Table 14

Percent Proficient on Several Composites at Various EOC Biology Test Cut Points

Cut Score on EOC Biology	Weighting by Equal Parts	Weighting by Testing Time	Weighting to Maximize Reliability
Percent Correct			
RS = 12	85.8	82.5	80.2
RS = 15	77.0	78.4	76.1
Expert Judgment			
RS = 18	72.1	67.1	72.4
RS = 20	66.8	60.6	72.4
RS = 22	58.9	49.8	67.1

Note. RS=raw score.

Discussion

Using the EOC tests instead of the general assessment subtests as accountability measures generally results in lower levels of proficiency for the group in this study, depending upon which cut score for the EOC is adopted. Composite scores that combine the two tests may be more appropriate measures of student achievement as more information is available from multiple types of assessments.

Composite score reliabilities calculated for a series of congeneric test parts showed that a simple appending of one test to the other resulted in a higher reliability than any other method of dividing the test into parts. This is likely because the end-of-course tests contain 30 items each and the mathematics and science subtest of the general assessment have 81 and 43 items, respectively, so the general assessment, the more reliable test, is weighted more in this case than the end-of-course test. The weights in the case

of the simple appending are the closest to the weights that turn out to maximize the reliability of the composite score.

Student proficiency was determined on three types of compensatory composite scores—with weights that were equal, that reflected the amount of time spent in testing, and that maximized reliability of the composite—and for various raw score cut points on the EOC test. Proficiency diminished as the EOC cut points increased, and there was a trend in the mathematics/Algebra I composites with proficiency increasing from equal weights, to weights based on testing time, to weights selected to maximize reliability. There was no similar trend in the science/Biology composites. While proficiencies based on the composite scores are not usually as high as those obtained when using the general assessment as the accountability measure, they may reflect a more appropriate measure of the construct because the composites are formed from two different assessments.

Comprehensive tests are aligned to academic standards but cover multiple years of instruction and classes of material. Rather than testing the knowledge accumulated over years of study, end-of-course tests assess what students learn in an individual course. Because the combination of a general assessment with an end-of-course test would measure both accumulated and course-specific knowledge in mathematics and science, end-of-course tests may prove a valuable addition for enhancing the information available for accountability decisions beyond what is offered by the general assessment. As states begin to struggle with measuring the very broad and inclusive standards found in the high school common core, accountability models will need to expand this work to determine the most appropriate methods for combining information from very different assessments, all measuring different aspects of broad content areas such as mathematics and science.

References

Achieve, Inc. (2007). *Aligned expectations? A closer look at college admissions and placement tests.* Washington, DC: Author.

Achieve, Inc. (2008). *American Diploma Project (ADP) end-of-course exams: 2008 annual report*. Washington, DC: Author.

Center on Education Policy. (2008). *State high school exit exams: A move toward end-of-course exams.* Washington, DC: Author.

Gewertz, C. (2007, May 16). States mull best way to assess their students for graduation. *Education Week*, pp. 1, 17.

Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp.65-110). Westport, CT: American Council on Education and Praeger.

Iowa Department of Education. (2009a). The Annual Condition of Education Report, 2009. Retrieved from http://publications.iowa.gov/9219

Iowa Department of Education. (2009b). Iowa Core Curriculum: K-12 Science. Retrieved from http://www.corecurriculum.iowa.gov/

Iowa Department of Education. (2010). Iowa Core Curriculum: K-12 Mathematics. Retrieved from http://www.corecurriculum.iowa.gov/

Kane, M., & Case, S. M. (2004). The reliability and validity of weighted composite scores. *Applied Measurement in Education, 17*(3), 221-240.

NGA Center for Best Practices. (2008). *Policies to improve instruction and learning in high schools*. Washington, DC: Author.

Rudner, L. M. (2001). Informed test component weighting. *Educational Measurement: Issues and Practice, 20*, 16-19.

Stiggins, R. (2006). Balanced assessment systems: redefining excellence in assessment. Educational Testing Service. Retrieved from http://www.assessmentisnt.com

Vranek, J.L. (2008). *The role of statewide end-of-course assessments in high school assessment systems: A study for the Washington State Board of Education.* Education First Consulting, LLC.