

**Comparing State High School Assessments to Standards for Success in
Entry-Level University Courses**

Richard S. Brown, Ph.D.

Assistant Professor

University of Southern California

213-740-9567

richarsb@usc.edu

David T. Conley, Ph.D.

Professor, University of Oregon

Director, Center for Educational Policy Research

541-346-6155

conley@uoregon.edu

RUNNING HEAD: COMPARING HIGH SCHOOL ASSESSMENTS

COMPARING HIGH SCHOOL ASSESSMENTS

Abstract

States have moved rapidly over the past twenty years to institute systems of standards and assessments. State assessments in particular take on added importance at the high school level as they are required for graduation by an increasing number of states. Federal legislation mandating testing in high school also serves to increase the stakes and impact of state exams. Many states are also using high school exams for postsecondary purposes, although the content and criterion validity of these exams in relation to students' post-high school pursuits is not well documented. While no state exam was developed with the express intent of aligning specifically with postsecondary education, it is nonetheless important to understand this linkage given the wide-ranging use of high school exams across the country. This study analyzes the content of state tests relative to a set of standards that identify knowledge and skills necessary for success in entry-level university courses. A total of 60 math and English assessments from 20 states were analyzed along a number of alignment dimensions. Exams were found to be moderately aligned with a subset of the university standards, but in an uneven fashion. English exams were somewhat more aligned than math exams, but math exams had high alignment in some specific standard areas, and English exams aligned poorly or not at all in areas requiring higher order thinking. In the future, states using high school exams for postsecondary purposes may want to examine the content of state standards and exams to determine their relationship to college readiness criteria.

COMPARING HIGH SCHOOL ASSESSMENTS

Introduction

The past two decades have witnessed the rise of state academic learning standards and assessments throughout the nation. The initial purpose of these standards and assessments was to establish high expectations to ensure all students were prepared to be economically productive citizens who were able to thrive and adapt in a rapidly changing world (Commission on the Skills of the American Work Force, 1990). However, an increasing number of states are beginning to consider the relationship between their exams and college admission, placement decisions, and financial aid. For example, Michigan chose to abandon its high school exam entirely, replace it with the ACT, and use the results for NCLB requirements as well as to determine the winners of state scholarships (Keller, 2005). Florida and Washington allow a designated score on the SAT to be substituted for the state high school exit exam or portions of it. Massachusetts is awarding scholarships to students who score at the highest performance level on the state high school exit exam. In addition, Illinois, Kentucky, and Colorado currently require all students to take the ACT. California students taking the California Standards Test (CST) in high school can access the California State University (CSU) Early Assessment Program by completing an extended version of the CST that yields placement information for CSU campuses (California State University System, 2006). The Texas Success Initiative expects higher education institutions to utilize information from the Texas Assessment of Knowledge and Skill (Texas Higher Education Coordinating Board, 2003). The City University of New York guarantees admission to students who obtain a designated level of performance on the New York Regents examination while in high school (City University of New York, 2006).

COMPARING HIGH SCHOOL ASSESSMENTS

Given the importance and influence of state high school assessments, a better understanding of the alignment between these tests and college readiness standards can offer insight into an issue of potentially fundamental importance, namely, the relationship between a high school education and college readiness. As the proportion of students attending college continues to increase and as college remediation rates remain stubbornly high, particularly at community colleges (National Center for Education Statistics, 2004), the signals that state high school exams send to students and teachers become increasingly important. Thus, a study investigating the relationship between state high school exams and college readiness standards may shed light on these two different conceptions of an education.

This study draws upon emerging theories of systems coherence and alignment (Clune, 1998; S. Fuhrman, 2001; S. H. E. Fuhrman, 1993; Goertz, Floden, & O'Day, 1996) to justify the examination of the relationship between state tests and postsecondary success standards. The basic premise of those theoretical frameworks is that, by creating more explicit connections between local educational systems and state standards and assessments, superior learning will result.

Signaling theory (Kirst & Venezia, 2004) also frames this analysis. As adapted to the high school-college transition process, this theory posits that high school students, teachers, administrators, and others receive signals from state standards and assessments and postsecondary admission requirements, among other sources, about what is important to teach and learn in high school. If the signals are unclear or contradictory, those who receive them cannot create programs or adapt practices that are internally consistent or that align with what comes next for students. When this is the case, the signal tends to be

misinterpreted or ignored. The potential power of the signal to the system is lost or greatly diminished.

State education policy has not necessarily been conceived with signaling as an explicit goal, but the ways in which administrators and teachers process policy signals appear to be a critical factor in determining the success of state education policies (Conley & Goldman, 2000; Spillane, 1998, 2000; Spillane & Callahan, 2000). The signals sent to schools become increasingly important as schools are subject to greater accountability through state and federal policies and as more students seek to pursue postsecondary learning.

This study does not assume that complete alignment or consistency among state assessments and college admission standards should be expected or that the lack of it is necessarily and automatically a bad thing. Rather, the basic premise of the study is that given the expanding use by states of high school examinations, it is worthwhile at the least to examine the degree of alignment that exists between state high school exams and college readiness standards.

Therefore, recalling Weick's (1976) well-known description of education as a "loosely coupled system," this study seeks evidence about the degree of alignment, or coupling, between the systems that may exist, either deliberately or inadvertently, as observed by the content of state high school exams in relationship to standards for success in entry-level college courses. If such testing is aligned with postsecondary success standards, it is reasonable to assume that schools that prepare students to do well on the test are also preparing students to do well in college. If alignment is low, it can be

COMPARING HIGH SCHOOL ASSESSMENTS

posited that high schools will be forced to choose between two competing sets of priorities, one set oriented to state exams and another focused on college readiness.

Whereas Weick's (1976) description of loosely coupled systems has been a very useful way to conceive of organizational relationships in public education, current policies adopted by the state and federal government call into question whether education will continue to function as a loosely coupled system, whether it will become more tightly coupled or whether the loose coupling will frustrate attempts at alignment. The adoption by states of "P-16" legislation (Blanco et al., 2003) is one more indicator that state policymakers are reconceptualizing the organizational structure of their public education systems from preschool through postsecondary education in ways that connect the levels more directly. Examining the relationship between state high school assessments and college readiness standards may help to ascertain the current relationship between levels in the educational system.

Methodology

The study applies and builds upon established alignment analysis methodology (Webb, 1997, 1999) to determine the congruence between the content of state high school assessments and a set of standards keyed to the knowledge and skills necessary for success in select American research universities. These standards, the most comprehensive of their type in the nation, are the Knowledge and Skills for University Success (KSUS) developed by the Standards for Success project (Conley, 2003). The KSUS standards were developed through the analysis of multiple data sources that indicated what was occurring in entry-level college courses. The study employed a convergent consensus process that gathered data at nine meetings in which over 400

COMPARING HIGH SCHOOL ASSESSMENTS

faculty members from Association of American University (AAU) universities participated. All taught entry-level courses. Documents from their courses were then subsequently analyzed and triangulated against information gathered via convergent consensus. The resulting standards were endorsed by 28 AAU universities as representative of what those universities expect of students in their entry-level courses. Subsequently, the standards have been utilized by the College Board as a foundational reference in reviews of its assessments and in the development of its own content standards in English and mathematics. The standards have also been adopted by a wide range of postsecondary institutions and at least one state higher education system as their official statement of expectations for incoming students.

This process resulted in comprehensive standards in six subject areas. For the purpose of this study, the KSUS standards in English and mathematics were employed as the representation of college readiness and, therefore, as the reference point against which state high school exams were analyzed. English standards are grouped under four headings: reading and comprehension, writing, critical thinking, and research skills. The KSUS mathematics standards comprise six topical areas: computation, algebra, geometry, math reasoning, trigonometry, and statistics.

This approach draws upon content alignment work that has taken place comparing state test content to state and national content standards (see, for example: Kendall & Marzano, 1995). Impara (2001) detailed several models for conducting content alignment studies, ranging from low to high levels of complexity. These methodologies generally employ content experts to rate the match between each test item's content and the content standards referenced in test specifications. The rating scales employed capture

COMPARING HIGH SCHOOL ASSESSMENTS

information on the breadth and the depth of content coverage. Each item can also be analyzed to rate its cognitive complexity. The extent of agreement among raters (rater reliability) is determined and presented along with the level of content match between the assessment and the content standards or test specifications and the appropriateness of item complexity.

A variation of this approach was utilized by Le, Hamilton, & Robyn (2000) in their alignment among secondary and post-secondary assessments in California. This approach is useful but limited in the amount of information that it provides. A more moderately complex approach described by LaMarca, et al. (2000) broadens the focus to include not only content match and depth of coverage, but also balance of coverage, performance match, and accessibility. Webb combined these various rating methods into one integrated system (Webb, 1997, 1999). Impara (2001) states, “because Webb’s conceptualization of alignment is comprehensive and extends far beyond the two previously described models, it can be adapted for use in virtually any context” (p. 4). As a result, this approach, with a few modifications, is the procedure used for determining the degree to which state assessments align with the KSUS standards.

While an argument can be made that the KSUS standards should be compared with state content standards, evidence suggests teachers orient their efforts more to state tests than to state standards (Koretz, Mitchell, Barron, & Keith, 1996; Lane, Stone, Parke, Hansen, & Cerillo, 1996; McDonnell & Choisser, 1997). The KSUS standards are derived as directly as possible from actual college classroom practices, not a theoretical knowledge domain that encompasses all “desirable” skills for incoming students. In this sense, comparing the KSUS standards to state assessments does connect two roughly

COMPARING HIGH SCHOOL ASSESSMENTS

parallel sets of knowledge expectations that are in operation and are connected to classroom instruction at their respective educational levels.

Alignment Rating Process

The rating process involved three stages. The raters first reviewed and scored each of the KSUS standards on a five-point scale adapted from Marzano (2001) to determine the depth of knowledge of each standard. The points of the scale were Retrieval, Comprehension, Analysis, Utilization, and Goal Setting/Monitoring. Marzano's scale is designed to reflect increasing cognitive complexity, more sophisticated uses of knowledge, and progressively higher levels of meta-cognitive functioning. The raters then used the five-point scale to assign a depth-of-knowledge rating to each assessment item within a discipline. Pilot study analysis of Depth of Knowledge ratings for the KUS indicates acceptable levels of consistency among the raters. These analyses utilized generalizability theory (Shavelson & Webb, 1991) to investigate the consistency of ratings. The generalizability coefficient (g-coefficient) is comparable to the more common reliability coefficient from classical test theory and indicates the percent of variability in the scores which are due to systematic influences. Estimated g-coefficients for the Mathematics KSUS ratings were lower than the estimated g-coefficients for the English/Language Arts KSUS. However, in both cases five to six raters were employed, resulting in reliability (generalizability) coefficients in excess of a desirable threshold of .80.

The third rating task required raters to review each assessment item to determine whether it addressed one or more of the KSUS objectives. Each assessment item was rated against each KSUS standard and a determination was made as to whether that item

COMPARING HIGH SCHOOL ASSESSMENTS

addressed each standard. In that way, an assessment item may be linked to any number of standards statements depending on its content coverage. The result of these three analyses was a matrix for each rater (KSUS standards by assessment items) that identified which high school test item addressed which KSUS standard, how many KSUS standards were addressed by how many state high school test items, and which KSUS standards were not addressed by any state high school test items. That is, for each standard, the number of assessment items addressing it, as well as the number of standards addressed by specific assessment items could be identified.

High school and university faculty were selected to serve as raters. All raters had experience with state high school assessments or direct involvement teaching entry-level university courses. Training consisted of practice sessions in which groups of ten items were rated and discussed. From these training sessions, scoring criteria and decision rules were developed and refined by the researchers, then applied consistently during actual rating sessions. No fewer than five raters rated each test. All tests were rated by mixtures of high school and university faculty.

Four Alignment Criteria and the Summary Alignment Index Value

This process of content focus alignment (Webb, 1997, 1999) yields scales to which criterion values are applied to reach overall conclusions about how well each assessment aligns with university standards. The four alignment criteria analyzed were Categorical Concurrence (the number of matches between KSUS objectives and individual assessment items), Depth of Knowledge (the cognitive complexity of test items in relation to KSUS objectives), Range of Knowledge (the spread of test items across KSUS objectives), and Balance of Representation (the ways in which test items

that matched a KSUS objective distributed onto KSUS objectives within a KSUS standard).

In addition to measures in these four areas, the study introduces a new metric for considering alignment: a single index-value of alignment that consolidates the multitude of alignment values for any given test. For example, in the area of mathematics, the alignment analysis using Webb's methodology generates values for the four alignment criteria just defined (Categorical Concurrence, Depth of Knowledge, Range of Knowledge, and Balance of Representation) for each of the six KSUS mathematics topical areas. In English/Language Arts, four alignment criteria for each of the four KSUS topical areas are provided. Thus, a total of twenty-four alignment measures for each state mathematics assessment and sixteen alignment measures in ELA result from the Webb approach. The Summary Alignment Index Value (SAIV) introduced here reduces the number of reported outcomes to a single SAIV score. Although we have retained and utilize the multiple criterion values in this study, we also reference the SAIV as an additional, shorthand way to understand relative alignment in the case where large numbers of tests are compared, as occurred in this study. We include it in this paper not as a fully proven metric but as a potentially useful tool that requires additional testing and refinement.

Categorical Concurrence

Raters first determined Categorical Concurrence, the match between the KSUS standards and objectives and each item from each state high school assessment. The goal was to determine two things: First, do state assessments cover areas deemed important to university success; and, second, are the KSUS standards and objectives in their current

COMPARING HIGH SCHOOL ASSESSMENTS

format useful for determining such relationships? Raters were asked to identify which (if any) KSUS standards and objectives were addressed by each assessment item. These ratings were then summarized across raters to determine the average number of items for each assessment addressing one or more KSUS standard. This criterion could range from 0 to the total number of items included on the examination. The recommended benchmark for this alignment criterion is that at least 6 items on average are aligned to each KSUS (Webb, 1999).

Depth of Knowledge Consistency

Each test item was analyzed to determine its Depth of Knowledge rating on the Marzano scale previously described. In addition, each KSUS objective was also given a Depth of Knowledge rating. State high school test items were then compared with their corresponding KSUS objective where categorical concurrence was established. If the assessment item was rated at the same or higher level of cognitive complexity as the KSUS objective with which it corresponded, a relationship was noted. Keeping with Webb's methodology, the Depth of Knowledge consistency is measured as the percentage of matches between an assessment item and a standard statement wherein the assessment item had an equivalent or higher Depth of Knowledge rating than the KSUS standard. Values for this criterion range from 0 to 1.0, with a recommended benchmark of $\geq .50$ (Webb, 1999). That is, for a state assessment to be considered adequately aligned with respect to Depth of Knowledge, at least half of the assessment items should be at or above the cognitive complexity level of the corresponding KSUS objective.

Range of Knowledge

COMPARING HIGH SCHOOL ASSESSMENTS

Range of Knowledge was determined by tallying the number of KSUS objectives that were addressed by one or more high school assessment item. This criterion gives an estimate of the breadth of knowledge addressed by the high school test. Webb describes the range-of-knowledge criterion as a means “to judge whether a comparable span of knowledge expected of students by a standard is the same as, or corresponds to, the span of knowledge that students need in order to correctly answer the assessment items/activities” (Webb, 1999, p. 8). This metric is created by calculating the percentage of objectives within a standard addressed by the assessment items for each rater. These rater estimates are then averaged across raters for a summary Range of Knowledge value. This criterion’s value ranges from 0 to 1.0, with a recommended benchmark of $\geq .50$. An assessment considered to be adequately aligned to a KSUS standard with respect to Range of Knowledge addresses at least half of the objectives within a KSUS standard. In this study, the number of objectives for a given standard varied from six to thirty in ELA (6 for Critical Thinking; 11 for Research Skills; 26 for Reading and Comprehension; and 30 for Writing) and from three to twenty-seven in mathematics (3 for Statistics; 4 for Trigonometry; 11 for Computation; 15 for Geometry; 24 for Algebra; and 27 for Mathematical Reasoning).

Balance of Representation

Balance of Representation identified how those items that matched specific KSUS objectives were distributed onto the KSUS objectives they matched. This criterion indicates the extent assessment items are evenly distributed across KSUS objectives. Like the Depth of Knowledge and Range of Knowledge criteria, Balance of Representation is expressed via an index score from 0 to 1.0, but has a recommended

COMPARING HIGH SCHOOL ASSESSMENTS

benchmark of .70 rather than .50. This higher threshold is established to ensure that reasonably good overlap exists between the test and the KSUS standards overall.

Summary Alignment Index Value (SAIV)

As noted earlier, each state assessment generates a multitude of alignment index values for the KSUS standards in English/Language Arts and Mathematics. While this information is useful for identifying how a state assessment aligns along four distinct dimensions of alignment, the complexity of this information makes it quite difficult to draw overall conclusions about the alignment of the assessment to the KSUS standards. A simpler, summary measure may be helpful in consolidating the information from as many as two dozen alignment measures for a given assessment, especially when comparisons across assessments are desirable.

To address this situation, we created the Summary Alignment Index Value (SAIV), which is the weighted average of the Range of Knowledge, Depth of Knowledge, and Balance of Representation indices for a given subject. Categorical Concurrence is not included in the summary calculations for two reasons. First, the Categorical Concurrence index is not comparable to the other indices, which range from 0-1. Rather, it is the total number of matches between state test items and KSUS objectives. Second, the Range, Depth, and Balance measures are calculated within a standard given the content matches identified without dependencies across standards. That is, it would be theoretically possible to have 100% Depth of Knowledge alignment for all standards in a subject area for a given assessment. However, given the nature of the content specificity of the standards, the Categorical Concurrence measure is not entirely independent across standards. That is, when a state test item addresses objectives

COMPARING HIGH SCHOOL ASSESSMENTS

within a KSUS standard, it is unlikely to address objectives within standards reflecting different content. No such interdependency across standards exists for the other alignment metrics.

To determine the SAIV for each state test, the alignment criterion values are analyzed to calculate the volume of the tetrahedron created by placing each of these values along one of three dimensions. Complete, or ideal, alignment (e.g., values of 1.0 on each of these three criteria) will yield a volume value of 1.0. Lesser values on each dimension will reduce the summary index value accordingly (see Figure 3 for a graphical representation of this idea). If we use the recommended benchmark values for alignment suggested by Webb (2002) (.50 for Depth of Knowledge and Range of Knowledge, and .70 for Balance of Representation), we calculate a benchmark SAIV of .563. That is, assessments yielding values below this point would be defined as not exhibiting adequate summary alignment for a standard, whereas those producing summary alignment values equal to or greater than this critical value would, by definition, be considered to exhibit adequate summary alignment for that standard.

--Insert Figure 1 about here --

An alignment index value is calculated for each standard within a subject area then combined using a weighted average (by number of objectives within each standard) over the standards to create a single summary value for each state assessment based on the prevalence of KSUS objectives per standard. These summary values (ranging from 0 to 1.0) are calculated for each state assessment and presented separately by subject area, thereby allowing for more interpretable comparisons across state assessments.

COMPARING HIGH SCHOOL ASSESSMENTS

Data Sources

Currently, 25 states have or will have by 2012 high school exams with implications for graduation (Center on Education Policy, 2006). These states enroll approximately 70% of public school students in the nation. All other states have some form of high school exam used to meet state and federal accountability requirements.

All states with high school examinations were invited to participate in the study. Sixty assessments, 30 in English/Language Arts and 30 in mathematics, from 20 states were ultimately analyzed. The number is greater than one per state because some states had multiple test forms, some had multiple end-of-course examinations, and some had separate reading and writing tests. Appendix A lists the states from which assessments were analyzed along with some important characteristics of the state's assessment system. In addition to multiple-choice tests, which every state had, 38 tests contained short or extended constructed response items, and two included performance tasks. Tests analyzed were those in effect in 2003. A review of participating states determined that eight have made some sort of change in their testing program during since 2003. Most changes involved new test forms or addition of a new assessment element such as end-of-course testing.

Results and Conclusions

Results are reported for Categorical Concurrence, Depth of Knowledge, Range of Knowledge, and Balance of Representation.

Categorical Concurrence

COMPARING HIGH SCHOOL ASSESSMENTS

On average, each KSUS standard had over 10 English/Language Arts (ELA) items addressing it for each of the state assessments ($M = 11$, $SD = 17$). However, looking at the average number of items across various ELA standards disguises certain trends. When viewed in terms of the frequency with which a test meets the alignment benchmark adopted for this study of 6 or more items for a given standard, only 42% of the assessments met the mark. That is, more than half of the time the state high school assessment did not have enough ELA items addressing a KSUS English standard to be considered adequately aligned.

Examining alignment in terms of each of the four KSUS standard areas (Reading Comprehension, Writing, Critical Thinking, and Research Skills) yields a more varied pattern. Reading Comprehension was adequately aligned 25 out of 30 times (83%) and writing half the time (15 of 30). Critical Thinking and Research Skills, however, were much less frequently aligned. Only 30% of the assessments (9 of 30) met the mark for Critical Thinking and only 1 (3%) did so for Research Skills.

The analysis of mathematics assessments yielded similar results. The average across all six KSUS math standards (Computation, Algebra, Math Reasoning, Geometry, Trigonometry, and Statistics) was more than 11 items per assessment addressing each standard ($M=11$, $SD=10$). Once again, averages varied across the KSUS standards. While overall categorical concurrence was higher for mathematics assessments than was found with the English/Language Arts assessments, two KSUS standards, Computation and Math Reasoning, accounted for most of the alignment. Math assessments met the minimum six items per standard threshold in Computation on 90% of the tests and in Math Reasoning on 93% of the tests. The other KSUS math standards were less

COMPARING HIGH SCHOOL ASSESSMENTS

frequently aligned. State tests met the alignment criterion for the Algebra standard 63% of the time (19 of 30); with Geometry 60% of the time (18 of 30); and with the Trigonometry standard only once among the 30 tests (3%). The KSUS Statistics standard was not adequately addressed by any state assessment.

Depth of Knowledge Consistency

ELA assessments align better in terms of depth of knowledge than categorical concurrence. More than 60% of the assessment items were at or above the cognitive complexity level of the corresponding KSUS objective for each KSUS English standard. While the level of cognitive complexity varied across KSUS standards, it varied less than it did for Categorical Concurrence. The cognitive complexity of state assessments was adequately aligned most frequently to Reading and Comprehension (93%), followed by Writing (70%), Research Skills (66%), then Critical Thinking (63%).

In the area of Mathematics, even more alignment with respect to Depth of Knowledge was found. On average almost 70% of the assessment item/objective comparisons found the cognitive complexity of the assessment items to be at or above that of the corresponding KSUS objective. Only the KSUS Trigonometry standard at 37% failed to exhibit high Depth of Knowledge Consistency. All other KSUS math standards demonstrated Depth of Knowledge alignment over 80% of the time.

Range of Knowledge

Range of Knowledge is more difficult to establish than Categorical Concurrence or Depth of Knowledge. To meet this criterion, the state assessment would have to address the majority of objectives within each KSUS standard. This task is made difficult by two issues. First, the state high school assessments are limited in the number

COMPARING HIGH SCHOOL ASSESSMENTS

of items they can contain due to time and other testing constraints. Second, the KSUS objectives within each KSUS standard are relatively extensive and comprehensive. English has a total of 73 objectives across four standards; Mathematics contains 84 objectives covering six standards. For this reason, this alignment criterion is less likely to be met than other alignment measures. In fact, this is exactly what we observed.

In ELA, on average, 29% of the objectives within a KSUS standard were addressed by state high school assessment items. Just over a quarter of the state high school tests (26%) met the threshold of addressing at least half of the KSUS objectives within a given standard. This finding was relatively consistent across standards. Writing was the only KSUS standard where half (15/30) of the state assessments addressed at least half of the KSUS objectives. Only 30% of the state assessments achieved adequate alignment on Range of Knowledge on the KSUS Reading and Comprehension standard; only 23% of state tests achieved adequate Range of Knowledge with the KSUS Critical Thinking standard. Despite the fact that the KSUS Research Skills standard had relatively few objectives, no state assessment met the benchmark for adequate range of Knowledge.

Mathematics, with its greater number of KSUS objectives, had even lower Range of Knowledge alignment. As in ELA, only about 28% of the objectives within the Mathematics KSUS standards were addressed. This resulted in less than 13% of the state mathematics assessments meeting the benchmark of addressing at least half of the KSUS objectives for a given KSUS Mathematics standard. This lack of Range of Knowledge alignment was consistent across all KSUS Mathematics standards. Only in the areas of Computation (23%) and Geometry (20%) did more than one in five state assessments

COMPARING HIGH SCHOOL ASSESSMENTS

address at least half of the objectives within these KSUS Mathematics standards. Only three did so in Algebra (10%), two in Trigonometry (7%), and only one did so for KSUS standards in Math Reasoning and Statistics.

Balance of Representation

Balance of Representation yielded higher results. This means that when assessments did address a given KSUS standard, they did so with a set of test items that were balanced within the concurrent KSUS standards. For ELA state assessments, the average balance index across all KSUS standards was .63 and in Mathematics it was .57. In ELA, 79% of state assessments demonstrated adequate balance of representation across all applicable KSUS standards. Similarly, 75% did so in Mathematics.

The KSUS standard of Reading and Comprehension received the highest rate of Balance of Representation at 97%. Other KSUS standards also showed high rates of Balance of Representation, with Writing 83%, Critical Thinking at 73.3%, and Research Skills at 63.3%.

It is worth reiterating that if Categorical Concurrence is low for a KSUS standard to begin with, the Balance of Representation statistic is less meaningful. In other words, of only a few items on a state test correspond with one or more KSUS objectives, the fact that they distribute in a balanced fashion is less significant.

In Mathematics, the Balance of Representation between state assessment items and KSUS standards of Computation, Algebra, Geometry, and Math Reasoning was high (90%). Only in the areas of Trigonometry (33%) and Statistics (0%) did the state assessments not demonstrate high Balance of Representation.

COMPARING HIGH SCHOOL ASSESSMENTS

Descriptive Statistics of SAIV

The SAIVs for these 30 state assessments in English/Language Arts and 30 in Mathematics are, on average, comparable across subject area. The average SAIV in ELA is .507, and in Mathematics it is .509 suggesting that the overall alignment across subject areas is comparable if one were to combine all the state assessments into one instrument. See Figures 2-3 below for the distributions of each summary index by subject area.

--Insert Figure 2 about here --

-- Insert Figure 3 about here --

While the means and standard deviations are comparable, the skewness differs between ELA and Mathematics.

Applying the derived benchmark for summary alignment, one-third of the state mathematics assessments demonstrates adequate summary alignment, while more than half (55%) of the state ELA assessments meet or exceed the SAIV of .563. Thus, one may conclude that with respect to this aspect of alignment, overall, these state assessments were more frequently in alignment with the Knowledge and Skills for University Success in English than they were in Mathematics. However, this information should be interpreted along with evidence of categorical concurrence alignment for a more complete picture of alignment between each test and the KSUS standards.

Interpretation of Findings

Alignment Differences between English and Mathematics

Range of knowledge was better in Mathematics than in English, in part because there are more assessment items on state mathematics tests than on English/Language Arts tests. The range limitation on English/Language Arts tests may reflect two

COMPARING HIGH SCHOOL ASSESSMENTS

phenomena: these tests often contain written passages students must read, and adequate time must be allotted for all students to read the passage, thereby reducing the overall number of questions that can be asked in the allotted time. Writing tests pose a different sort of problem because they often consist of a single prompt upon which students must write. Even when the scoring guides are included in the content analysis, it is still difficult to make very many connections between the single writing prompt and scoring guide and many of the KSUS standards.

Another partial explanation is that the university standards and objectives in Reading Comprehension and Writing are more demanding and detailed than are corresponding state assessment items. The university standards expect students to use variety in sentence structure; demonstrate mastery of grammar and word usage; show evidence of a controlled style and voice; know the difference between a topic and a thesis; use a variety of methods to develop arguments, including compare-contrast, inductive-deductive, and general-specific; and distinguish between formal and informal styles, among other skills. State tests rarely cover these areas in adequate depth, nor do state writing tasks allow students to demonstrate many of these skills. Scoring guides used to score writing assessments may reference some of these objectives, but must omit many others.

State assessments do not align as well with the KSUS standards in Trigonometry and Statistics for Mathematics. These suggest two separate issues. Many state mathematics tests are administered at the tenth grade, a time when many students have not yet been exposed to trigonometry or statistics, and many states may have chosen not to include , trigonometry and statistics in their state standards.

COMPARING HIGH SCHOOL ASSESSMENTS

In English, where alignment is poorest with Critical Thinking and Research Skills, the problem is somewhat different, although here, too, it is entirely likely that state standards simply do not include specific expectations for proficiency in critical thinking and research skills. The types of skills identified by the KSUS standards in these topical areas are more challenging to assess via standardized testing forms, which rely on multiple-choice question formats. Although critical thinking of a sort can be discerned via multiple-choice tests, more complex cognitive tasks are generally required to ascertain students' abilities in this area more accurately. Similarly, paper-and-pencil tests do not assess well research skills, nor are these skills necessarily viewed as the primary domain of English/Language Arts instruction in many high schools.

For most states, Depth of Knowledge consistency between KSUS standards and state assessment items reaches sufficient levels of alignment. This is especially the case for almost all areas of mathematics, with Trigonometry being the sole exception. This can be explained at least in part based on the nature of the rating scale. Mathematics problems tend to fall into a more restricted range on the scale, and there are many KSUS standards in mathematics, including those that cover more basic math concepts. It becomes relatively easier to match problems with KSUS standards due at least in part to these factors. This reflects the fact that the KSUS standards are inclusive of all the prerequisite knowledge needed for success in entry-level math courses and not simply a listing of the college-level skills. Inclusion of basic computation and reasoning skills are a reflection of the fact that university faculty observed that entering students were often lacking in basic math skills.

COMPARING HIGH SCHOOL ASSESSMENTS

For most state assessments, Range of Knowledge alignment falls below acceptable benchmark levels. This is likely due at least in part to the fact that there are more KSUS objectives than assessment items in both Mathematics and English. This finding suggests that as states revise their assessments (and the standards that underlie them), it may be within the grasp of their assessment system to provide useful diagnostic feedback about college readiness to students by adding or redistributing items.

Considering the Overall Relationship between State Assessments and KSUS standards

It appears that most state assessments are not well aligned *overall* to the knowledge and skills for university success in Mathematics, but most are better aligned in English. The alignment in Mathematics is concentrated largely within the KSUS standards of Computation and Mathematical Reasoning. This is largely a function of the inability of state exams to cover the scope of the college readiness standards, which encompass a range of skills that can best be demonstrated only through classroom-based tasks, assignments, and tests combined with the fact that state tests are given in grades 10 and 11 predominantly, which means the tests cannot measure everything that students learn throughout high school. Mathematics might have been expected to fare somewhat better, particularly in the area of problem solving, given the broad reach of the standards developed by the National Council of Teachers of Mathematics (1989) and their influence on the development of math standards in many states.

However, another consideration is the fact that state standards-based tests tend to concentrate item difficulties around the cut score and generate much less information at other points along the ability continuum, which may restrict the skills assessed to those that cluster in that part of the distribution. In other words, many state assessments test

COMPARING HIGH SCHOOL ASSESSMENTS

fewer skills with greater redundancy within a specified range of difficulty. This strategy is useful for making a determination of where a student falls in relation to a cut score, but is not so useful if the goal is to ascertain performance along a wider cognitive continuum that extends to readiness for postsecondary education.

Conclusion

The study concludes that state high school assessments and the knowledge and skills necessary for university readiness align in areas that might be characterized as more basic and do not align as well in areas requiring more sophisticated cognitive functioning. The conclusion is that high school exams cover only a portion of what is necessary for college readiness and that, from a criterion validity perspective, the results from these tests should be interpreted with caution if used for post-secondary decisions. This finding is particularly important in light of recent initiatives by some states to consider linking their current exams to placement, financial aid, and admission decisions.

The alignment between state assessments, which are most commonly given in tenth or eleventh grade, and university readiness standards is better than might have been expected in some areas, given that state tests were not designed with postsecondary standards as a reference point. This observation reaffirms the common-sense conclusion that some reasonable degree of alignment already exists between a high school education in most parts of the country and at least a subset of college readiness skills. The more pressing question is whether this baseline level of alignment is sufficient to result in well prepared students at a time when the proportion of high school graduates pursuing postsecondary education is increasing.

COMPARING HIGH SCHOOL ASSESSMENTS

However, as noted, this alignment is not evenly distributed across the standards. Reading and writing standards in English and computation and math reasoning in Mathematics explain most of the alignment. Other standards areas such as research skills, critical thinking, statistics, and trigonometry are seriously underrepresented or non-existent in state tests.

If states do wish to employ their high school exams to generate information on college readiness or placement, they will likely need to revisit the content domains from which exam items are drawn, the number and difficulty of test items, and the formats used for testing. The California State University Early Assessment Program provides one model of how to expand a state test so that the results are useful for college placement purposes while simultaneously providing cues to high school students and teachers on what should be occurring in high school classes for more students to be college ready.

A more fundamental issue is whether high school tests should have any relation to college readiness. This is perhaps a more philosophical consideration left to the deliberations of educators, policy makers, students, and parents. Given the range of current efforts and initiatives to redesign high schools fundamentally, a goal to which many states have subscribed publicly (American Diploma Project, 2006), the issue of the purpose of high school is likely to be examined and resolved through the actions of the key constituents in the educational system.

This redesign process could be facilitated to a degree by better alignment between high school and college. Greater alignment has proven challenging but not impossible. Programs such as Advanced Placement[®] and International Baccalaureate have devised assessment methods that connect directly to college readiness and also include

COMPARING HIGH SCHOOL ASSESSMENTS

performance tasks to measure complex cognitive processes, but these systems are more costly and complex to construct and maintain, and currently are restricted to only the highest achieving high school students. Oregon's Proficiency-based Admission Standards System (PASS) makes provision for a range of assessment scores in combination with collections of evidence from classroom-based student work to be used to meet admission requirements at state universities (Conley, 2004).

Connecting high school assessments with college readiness may prove impractical or even undesirable in many states. However, this study reinforces the standard caveat that states should take care not to use exam results for purposes for which the exam was not designed. This observation takes on added importance during a period when policy makers are seeking to redesign or restructure educational systems so that more students attend college.

This line of research and its early findings help frame the question of the relationship between high school exams and college readiness in order to set the stage for more focused studies of state high school assessments. Future studies need to look much more closely at the means by which the content to be tested on state high school exams is selected and on the reasonable inferences and conclusions that can be drawn from scores on these tests. In those cases where high school exams are used for post-secondary purposes, state assessment programs should be studied carefully to determine whether their content does align with college readiness standards and whether that alignment is limited to less cognitively complex aspects of college readiness.

This study offers initial insights into the relationship between state tests and university readiness standards and suggests that although a degree of alignment exists

COMPARING HIGH SCHOOL ASSESSMENTS

with a specific subset of all university readiness standards, this alignment is probably not adequate to consider state high school examinations in their current manifestations as significant sources of information on the full range of knowledge and skills necessary for postsecondary readiness. Given the scope of college readiness standards and their cognitive complexity, it is unlikely state assessments alone will ever capture information on the full range of content knowledge and cognitive skills, or “habits of mind,” essential to college success.

However, macro policy trends suggest that policy makers and their constituents will eventually demand greater alignment between state high school exams and college as an ever-increasing proportion of students choose to attend postsecondary education. A well formulated program of research to establish the validity of state exams for multiple purposes can help “second generation” state assessments achieve clearer goals and provide better information than the first-generation exams instituted over the past twenty years.

COMPARING HIGH SCHOOL ASSESSMENTS

References

- American Diploma Project. (2006). *Closing the expectations gap 2006: An annual 50-state progress report on the alignment of high school policies with the demands of college and work*. Washington, DC: Achieve, Inc.
- Blanco, C., Crowe, E., Lingenfelter, P. E., Longanecker, D. A., L'Orange, H. P., Rainwater, T., et al. (2003). *Student success: Statewide P-16 systems*. Colorado: State Higher Education Executive Officers.
- California State University System. (2006). Early Assessment Program. Retrieved August 24, 2006, from <http://www.calstate.edu/eap/>
- Center on Education Policy. (2006). *State high school exit exams: A challenging year*. Washington, DC: Author.
- City University of New York. (2006). *Admission guide- CUNY*. New York: Author.
- Clune, W. (1998). *Toward a theory of systemic reform: The case of nine NSF statewide systemic initiatives. Research Monograph No. 16*. Madison WI: National Institute for Science Education.
- Commission on the Skills of the American Work Force. (1990). *America's choice: High skills or low wages!* Rochester, NY: National Center on Education and the Economy.
- Conley, D. T. (2003). *Understanding university success*. Eugene, Oregon: Center for Educational Policy Research, University of Oregon.
- Conley, D. T. (2004). Proficiency-based admissions. In W. J. Camara & E. W. Kimmel (Eds.), *Choosing students: Higher education tools for the 21st century*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

COMPARING HIGH SCHOOL ASSESSMENTS

- Conley, D. T., & Goldman, P. (2000). Half full, half empty? Educator response over time to state-level systemic reform initiatives. *International Journal of Educational Reform, 9*(3), 249-269.
- Fuhrman, S. (Ed.). (2001). *From the capitol to the classroom: Standards-based reform in the states*. Chicago: The University of Chicago Press.
- Fuhrman, S. H. E. (1993). *Designing coherent education policy: Improving the system*: Consortium for Policy Research in Education, New Brunswick, NJ.
- Goertz, M. E., Floden, R., & O'Day, J. (1996). *The bumpy road to education reform. CPRE policy briefs*. Philadelphia, PA.: Consortium for Policy Research in Education.
- Impara, J. C. (2001). *Alignment: One element of an assessment's instructional utility*. Paper presented at the National Council on Measurement in Education, Seattle, WA.
- Keller, B. (2005). Michigan chooses ACT as high school exam. *Education Week, 25*(6), 20.
- Kendall, J. S., & Marzano, R. J. (1995). *The systematic identification and articulation of content standards and benchmarks. Update*. Aurora, Colorado: Mid-Continent Regional Educational Lab.
- Kirst, M. W., & Venezia, A. (2004). *From high school to college: Improving opportunities for success in postsecondary education*. San Francisco: Jossey-Bass.
- Koretz, D. M., Mitchell, K. J., Barron, S., & Keith. (1996). *Perceived effects of the Maryland state assessment program* (No. CSE Technical Report #406). Los

COMPARING HIGH SCHOOL ASSESSMENTS

- Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, Los Angeles.
- La Marca, P., Redfield, D., Winter, P., Bailey, A., & Despriet, L. (2000). *State standards and state assessment systems: A guide to alignment*. Washington, D.C.: Council of Chief State School Officers.
- Lane, S., Stone, C., Parke, C., Hansen, M., & Cerillo, T. (1996). *Consequential evidence for MSPAP from the teacher, principal, and student perspective*. Paper presented at the National Council on Measurement in Education, New Orleans, Louisiana.
- Le, V., Hamilton, L., & Robyn, A. (2000). Alignment among secondary and post-secondary assessments in California. In *Crucial Issues in California Education 2000*. Berkeley, CA: Policy Analysis for California Education.
- Marzano, R. J. (2001). *Designing a new taxonomy of educational objectives. Experts in Assessment*. Thousand Oaks, California: Sage.
- McDonnell, L. M., & Choisser, C. (1997). *Testing and teaching: Local implementation of new state assessments* (No. CSE Technical Report #442). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles.
- National Center for Education Statistics. (2004). *Conditions of education 2004*. Washington, D.C.: U. S. Department of Education.
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

COMPARING HIGH SCHOOL ASSESSMENTS

- Spillane, J. P. (1998). State Policy and the non-monolithic nature of the local school district: Organizational and professional considerations. *American Educational Research Journal*, 35(1), 33-63.
- Spillane, J. P. (2000). Cognition and policy implementation: District policymakers and the reform of mathematics education. *Cognition and Instruction*, 18(2), 141-179.
- Spillane, J. P., & Callahan, K. A. (2000). Implementing state standards for science education: What district policy makers make of the hoopla. *Journal of Research in Science Teaching*, 37(5), 401-425.
- Texas Higher Education Coordinating Board. (2003). *Subchapter C. Texas Success Initiative*. Austin, Texas: Author.
- Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education* (Research Monograph No. 6). Washington, D.C.: Council of Chief State School Officers.
- Webb, N. L. (1999). *Alignment of science and mathematics standards and assessments in four states* (Research Monograph No. 18). Madison, Wisconsin: Wisconsin Center for Education Research.
- Webb, N. L. (2002). *An analysis of the alignment between mathematics standards and assessments for three states*. Paper presented at the American Educational Research Association, New Orleans, LA.
- Weick, K. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, 21(2), 1-19.

COMPARING HIGH SCHOOL ASSESSMENTS

Appendix A. Characteristics of State Exams Analyzed

State	Test ¹	Test Name	Grade level given	Type ²	Grad Req. (y/n)	Test Developer
Colorado	Math	Colorado Student Assessment Program- Mathematics	9th	1, 2, 3, 4	No	CTB-McGraw Hill
	English	Colorado Student Assessment Program- Reading & Writing	9th	1, 2, 3, 4	No	CTB-McGraw Hill
Connecticut	Math	Connecticut Academic Performance Test-2 (CAPT-2)- Mathematics	10th	1, 3, 4, 7	No	Harcourt Educational Measurement
	English	Connecticut Academic Performance Test-2 (CAPT-2)- Language Arts, Reading, Writing	10th	1, 3, 4, 7	No	Harcourt Educational Measurement
Illinois	Math	Prairie State Achievement Examination- Mathematics	11 th	1	Yes	ACT
	English	Prairie State Achievement Examination- Writing & Reading	11 th	1	Yes	ACT
Kentucky	Math	Kentucky Core Content Tests (KCCT)- Mathematics	10 th	1, 4	No	CTB-McGraw Hill
	Math	Terra Nova (CTBS/5)- Mathematics	9 th	1	No	CTB-McGraw Hill
	English	Kentucky Core Content Tests (KCCT)- Reading & Writing	10 th	1, 4	No	CTB-McGraw Hill
	English	Writing Portfolio	12 th	6	No	CTB-McGraw Hill
	English	Terra Nova (CTBS/5)- Language Arts, English & Reading	9 th	1	No	CTB-McGraw Hill
Maine	Math	Maine Educational Assessment- Mathematics	11 th	1, 3, 4	No	Measured Progress
	English	Maine Educational Assessment- Language Arts, Reading & Writing	11 th	1, 3, 4	No	Measured Progress
Massachusetts	Math	Massachusetts Comprehensive Assessment System (MCAS)- Mathematics	10 th	1, 3, 4	Yes	Harcourt Educational Measurement
	English	Massachusetts Comprehensive Assessment System (MCAS)- Language Arts & Writing	10 th	1, 3, 4	Yes	Harcourt Educational Measurement
Michigan	Math	Michigan Educational Assessment Program (MEAP)- Mathematics	11 th	1, 3, 4	No	Riverside Publishing
	English	Michigan Educational Assessment Program (MEAP)- Reading & Writing	11 th	1, 3, 4	No	Riverside Publishing
Minnesota	Math	Minnesota Comprehensive Assessment- Mathematics	10 th	1, 3	No	Internally Developed
	English	Minnesota Comprehensive Assessment- Reading	10 th	1, 3	No	Internally Developed
	English	Basic Standards Writing	10 th	3	Yes	National Computer Systems
Mississippi	Math	Functional Literacy Exam- Mathematics	11 th	1, 4	Yes	National Computer Systems
	Math	Subject Area Testing Program- Mathematics	EOC ³	1, 3, 4	Yes	CTB-McGraw Hill
	English	Functional Literacy Exam- Reading & Writing	11 th	1, 4	Yes	National Computer Systems
	English	Subject Area Testing Program- Language Arts	EOC ³	1, 3, 4	Yes	CTB-McGraw Hill
Missouri	Math	Missouri Assessment Program- Mathematics	9 th	1, 3, 4	No	CTB-McGraw Hill
New Hampshire	Math	New Hampshire Educational	10 th	1, 4	No	Measured Progress

COMPARING HIGH SCHOOL ASSESSMENTS

Hampshire		Improvement and Assessment Program- Mathematics				
	English	New Hampshire Educational Improvement and Assessment Program- Language Arts	10 th	1, 4	No	Measured Progress
New Jersey	Math	New Jersey High School Proficiency Assessment (HSPA)- Mathematics	11 th	1, 3, 4	Yes	McDougal Little
	English	New Jersey High School Proficiency Assessment (HSPA)- Language Arts	11 th	1, 3, 4	Yes	McDougal Little
New York	Math	Regents Comprehensive Examination Program- Mathematics	EOC ³	1, 3, 4	Yes	Internally Developed
	English	Regents Comprehensive Examination Program- Language Arts	EOC ³	1, 3, 4	Yes	Internally Developed
Oregon	Math	Oregon Statewide Assessment- Mathematics	10 th	1, 4	Yes	SEA
	English	Oregon Statewide Assessment- Language Arts & Writing	10 th	1, 4	Yes	SEA
Pennsylvania	Math	Math, Reading Assessment- Mathematics	11 th	1, 2, 3, 4	No	Data Recognition Corporation
	English	Math, Reading Assessment- Reading & Writing	11 th	1, 2, 3, 4	No	Data Recognition Corporation
Texas	Math	Texas Assessment of Knowledge and Skills- Mathematics	11 th	1, 3, 4	Yes	Harcourt Educational Measurement
	English	Texas Assessment of Knowledge and Skills- Language Arts	11 th	1, 3, 4	Yes	Harcourt Educational Measurement
Utah	Math	Utah Core Criterion Referenced Tests- Mathematics	9 th	1	No	IBRIC
	Math	Stanford Achievement Test, 9- Mathematics	11 th	1	No	Harcourt Educational Measurement
	English	Stanford Achievement Test, 9- Language Arts & Reading	11 th	1	No	Harcourt Educational Measurement
	English	Utah Core Criterion- Referenced Tests- Language Arts	9 th	1	No	IBRIC
Virginia	Math	Stanford Achievement Test, 9- Mathematics	9 th	1	No	Harcourt Educational Measurement
	Math	Standards of Learning Assessment (SOL Tests)- Mathematics	EOC ³	1, 4	Yes	Harcourt Educational Measurement
	English	Standards of Learning Assessment (SOL Tests)- Reading & Writing	EOC ³	1, 4	Yes	Harcourt Educational Measurement
	English	Stanford Achievement Test, 9- Language Arts & Reading	9 th	1	No	Harcourt Educational Measurement
Washington	Math	Iowa Test of Educational Development (ITED)- Mathematics	9 th & 11 th	1	No	Riverside Publishing
	Math	Washington Assessment of Student Learning- Mathematics	10 th	1, 3, 4	No	Riverside Publishing
	English	Iowa Test of Educational Development (ITED)- Language Arts & Reading	9 th & 11 th	1	No	Riverside Publishing
	English	Washington Assessment of Student Learning- Reading & Writing	10 th	1, 3, 4	No	Riverside Publishing
Wyoming	Math	Wyoming Comprehensive Assessment System- Mathematics	11 th	1, 3, 4	No	CTB McGraw-Hill
	English	Wyoming Comprehensive Assessment System- Reading & Writing	11 th	1, 3, 4	No	CTB McGraw-Hill

COMPARING HIGH SCHOOL ASSESSMENTS

¹ Does not total 60 due to multiple forms and multiple course-specific tests in some states

²Test Types:

1. Multiple Choice
2. Fill in the Blanks
3. Short Constructed Response
4. Extended Constructed Response
5. Observation
6. Examples of Student Work/ Portfolio
7. Task Performance
8. Projects, Exhibitions, Demonstrations
9. None Available

EOC³: End of Course

Figure 1. Summary Alignment Index Tetrahedron

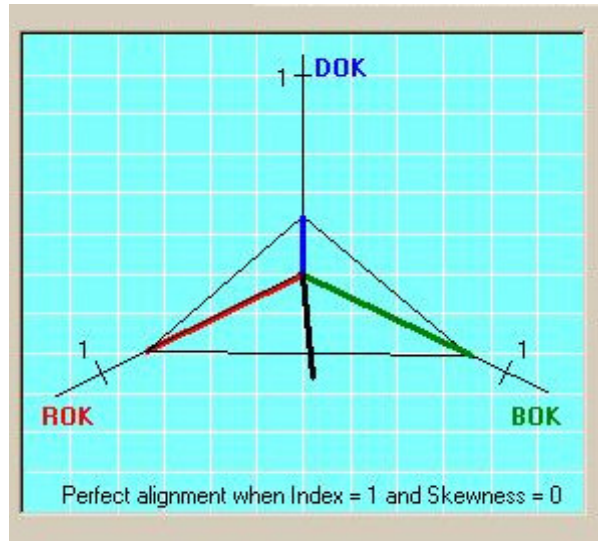
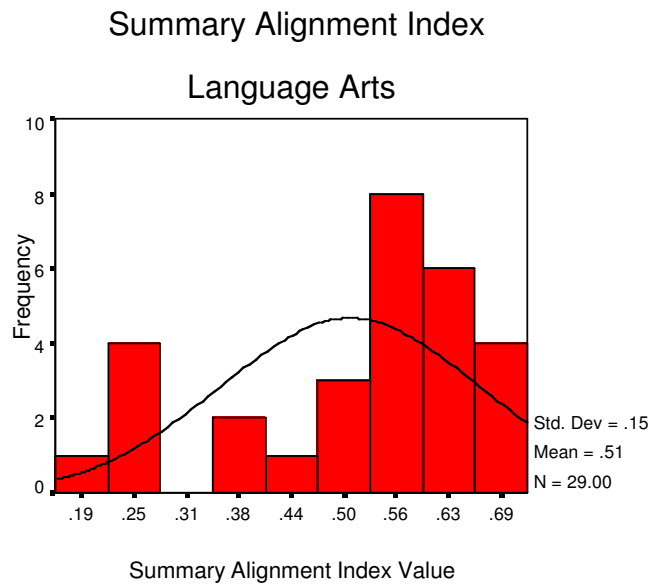


Figure 2: Summary Alignment Index Value Distribution: English/Language Arts



Alignment Summary Index

Mathematics

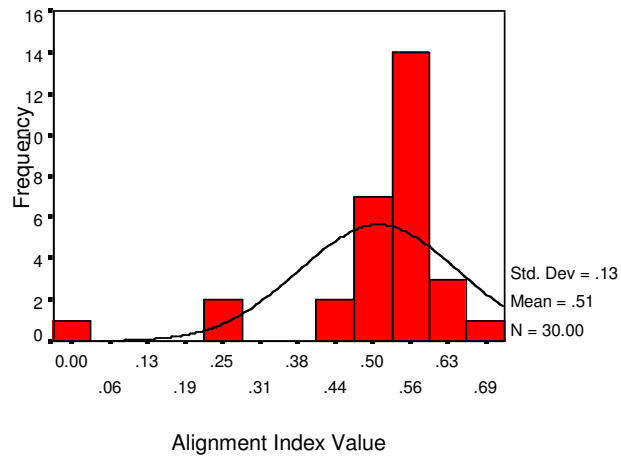


Figure 3: Summary Alignment Index Value Distribution: Mathematics