Chapter 1
Introduction

Background: The Current Role of Assessments in Secondary and Postsecondary Schooling

The high rate of remediation at the postsecondary level has attracted considerable attention in recent years. Nationally, nearly three out of ten first-time college students enroll in at least one remedial writing, reading, or mathematics course (National Center for Education Statistics (NCES), 1996). This has prompted concern from many policymakers and educators because students requiring remediation are less likely to earn a degree, especially if more than one remedial course is needed (NCES, 1997). Adelman (1998) found that students who took two or more remedial courses were nearly 25% less likely to earn a bachelor’s degree than students who did not take any remedial courses. Remediation also increases length of undergraduate studies for students, as well as costs for higher education institutions. A recent study found that community colleges in Michigan alone spent more than $65 million annually on remedial instruction (Greene, 2000), and estimates of the annual budgetary costs of remedial programs to the nation’s postsecondary institutions have reached as high as one billion dollars (Breneman, 1998).

A potential cause of remediation is that students do not have a clear sense of the competencies required for non-remedial, college-level courses (Center for an Urban Future, 2001; Education Trust, 1999, 2000; Kirst, 1997, 1998a, 1998b, 1999; National Commission on the High School Senior Year, 2001; Rodriguez, 1995). Differences among the myriad of assessments that prospective college students take are thought to send a confusing array of signals about the competencies needed for postsecondary work (Education Trust, 1999; Kirst, 1997). At the secondary level, almost all students take a state achievement test that measures student progress toward state standards (American Federation of Teachers [AFT], 2001). In addition to the state achievement test, college-bound students usually take another bevy of assessments, ranging from nationally administered college admissions exams to college-specific course placement tests. Prospective college students may also take end-of-course exams, which assess student knowledge of a particular course. Because state achievement, college admissions,
college placement, and end-of-course exams vary in content, format, and scope (Education Trust, 1999), students may receive inconsistent messages about the kinds of skills and knowledge they should master in order to enroll in non-remedial, college-level courses. If the tests were more aligned with one another (i.e., measured similar kinds of skills and knowledge), students would receive more uniform signals about the preparation needed to be ready for college-level work (Kirst, 1997).

Differences in test use necessarily limit the extent to which tests can be aligned. Ideally, greater alignment among state achievement, college admissions, college placement, and end-of-course exams should be achieved without compromising the usefulness of those measures. When tests are designed for different uses, test content can be expected to vary somewhat.\footnote{Differences in content standards or test specifications also contribute to variations in test content. However, a systematic analysis of content standards or test specifications for each test is beyond the scope of this study.} To illustrate, consider a state achievement test and a college admissions exam. A state achievement test is intended to monitor student achievement toward state standards, and is taken by students of all proficiency levels, including the approximately 30 percent who do not plan to attend college (NCES, 1997). In contrast, a college admissions test is intended to sort students more qualified for college-level work from those less qualified, and is taken by a self-selected group of higher-achieving students interested in attending a four-year postsecondary institution. Given the differences in purpose, it is reasonable to expect that a college admissions exam will include more difficult items than a state achievement exam. Modifying the content of a state achievement test so that it is closely aligned to that of a college admissions exam may undermine the utility of the state achievement test as an indicator of knowledge of the general student population statewide.

The Goal of this Study

The primary goal of this study is to investigate alignment among different types of tests in five case study states. We compare state achievement, college admissions, college placement, and end-of-course exams, classifying items along several dimensions. As a result, we obtain for each state a summary of the ways in which the assessments are and are not aligned with one another, and discuss possible implications. In particular, we
consider the signaling functions of the tests and discuss instances in which discrepancies among exams are warranted.

This analysis is part of a larger study, the Stanford Institute for Higher Education Research's Bridge Project, which is examining the transition from secondary to postsecondary education. The Bridge Project is exploring ways in which to improve opportunities for all students to enter and succeed in higher education by strengthening the alignment between higher education admissions requirements and secondary education curriculum, content standards, and assessments. The importance of having a more coordinated secondary and postsecondary system is underscored by a 1995 National Education Association (NEA) survey with 58 house and senate education committee chairs in 49 states (NEA, 1995). Eighty-two percent of the chairs polled indicated that improving the connections between colleges and secondary schools were among their highest priorities for higher education.

Structure of the Present Report

There are five states serving as the case sites for the present report: California, Georgia, Maryland, Oregon, and Texas. We begin by presenting the conceptual framework that guides our analysis; namely, we discuss the different purposes of the tests we analyzed, focusing on how variations in test functions affect the degree of alignment. We then discuss the merits of greater alignment among different assessments. We conclude by defining what we mean when we refer to tests as “misaligned.” Chapter 2 discusses the study’s limitations and methodology, focusing on the classification systems used to code the tests. In Chapters 3-7, we provide the results for each state, presented by chapter. For each state, we provide a description of the testing environment, the assessments specific to the particular state, and the results of the alignment analyses. Discrepancies within tests of similar purposes, as well as discrepancies across tests of different purposes, will be presented. Because the case reports for each site are designed to be parallel, there will be substantial overlap among Chapters 3-7. We conclude with

---

2 Wisconsin was not represented in the study.
3 Although Illinois is one of the sites selected as part of the larger Bridge Project, it is not included in the present study because items from its newly developed state achievement exam were not available.
Chapter 8, which provides an overall summary of alignment across the five case study sites, and discusses implications of our findings.

A Conceptual Framework for Analyzing Alignment Among Different Types of Tests

In this section, we discuss four types of assessments examined in this study: state achievement, college admissions, college placement, and end-of-course exams. We describe the primary functions of each type of test, and discuss how the content of each type of test reflects its purpose. We also highlight discrepancies that can be expected a priori among different types of assessments.

Tests with Different Functions

State Achievement Tests

In this study, state achievement tests refer to any assessment that is used by the state to monitor student achievement statewide and/or to track student performance in relation to state standards. In the past, these were the primary functions of state achievement tests (Feuer, 1992), but recently state achievement exams have taken on other functions as well. State achievement tests are seen as a potential lever for instructional or curricular reform (Popham, Cruse, Sandifer, and Williams, 1985; Shepard, 2000), and they have been used to reward and sanction schools, as well as to grant students a high school diploma (National Research Council [NRC], 1999).

The content of many state achievement tests is guided by each state’s standards, which specify the skills and knowledge students are expected to learn. Most state achievement tests are designed to provide a broad survey of student proficiency, although a few states have moved away from survey assessments towards course-based assessments. Despite the recent movement towards more rigorous test content (see Achieve, 1999), some state achievement exams, particularly those required as a

---

4 It is important to keep in mind that these categorizations are not mutually exclusive, but represent one method of organizing different kinds of tests.
5 We have excluded any tests intended solely for bilingual or disabled students.
6 According to a report by the AFT (2001), all states have developed standards for math and English at the high-school level.
prerequisite for graduation, continue to emphasize content from basic courses (Education Trust, 1999). The reasons for this will be explained below.

Many state achievement tests support criterion-referenced or standards-based interpretations, which means that the performance of a student is reported relative to a standard or criterion of performance, not to the performance of others. Although accurate categorization of students into the proper performance category (e.g., below basic, basic, proficient, advanced) is always a concern, classification accuracy becomes crucial if scores on a test are used as the basis for awarding or denying students their high school degree. Because the expected passing rate on most graduation tests is typically high (Bolon, 2000), the tests need to discriminate within the lower region of the achievement distribution. To maximally discriminate within this region, there needs to be many items assessing lower-level skills and material from basic courses. Basic content is also included on state achievement tests to accommodate the wide range of student proficiency levels within the state, including those at the lower-end of the achievement distribution.

State achievement tests that are used to award students their high school diploma must withstand legal challenges. By law, students cannot be held accountable for content that they have yet had an opportunity to learn. Therefore, content on graduation tests must mirror the content taught in most classrooms. Concerns about the legal defensibility of graduation exams may dictate that textbook-like items comprise the majority of the test at the expense of non-routine, problem-solving and inference items.

In response to recent demands that state achievement exams be “tests worth teaching to,” state achievement tests may also include some open-ended items. Open-ended items are thought to better foster problem-solving skills than are multiple-choice items (Resnick & Resnick, 1992), and are also viewed as a better means of influencing classroom instruction than are multiple-choice questions (Ruddell, 1985). Multiple-choice questions, however, remain by far the predominant format on most state achievement tests.

---

7 Readers should keep in mind that all graduation tests are considered state achievement tests, but not all state achievement tests are graduation tests.
College Admissions Tests

Because courses and grading standards of different high schools have been shown to vary widely (Office of Educational Research and Improvement, 1994), college admissions exams can enable fair comparisons among applicants by providing higher education institutions with a common measure in which students can be evaluated. Most college admissions tests favor the multiple-choice format because it is less costly, more reliable, and are perceived to be more objective than open-ended formats (Rothman, 1995). College admissions measures are used nationwide, so they do not reflect the content standards of any particular state, but instead assess general high school curricula. That is, they measure student knowledge of content from various courses found in most high schools. The SAT I is a notable exception among college admissions measures. Although some high school curriculum is included, it is not intended to measure mastery of high-school curricula (Wightman & Jaeger, 1988). Instead, it is a measure of general reasoning skills developed over years of schooling. More details about the SAT I will be provided in Chapter 2.

College entrance tests provide norm-referenced interpretations, which means that the performance of a student is reported relative to the performance of others. This method of reporting is consistent with the primary function of college admissions exams, which is to sort applicants more qualified for college-level work from those less qualified (Feuer, 1992). Because the population of examinees taking college admissions tests is of higher proficiency, college admissions tests need to discriminate among the higher regions of the achievement distribution. Therefore, an emphasis on problem solving and advanced content is appropriate.

College Placement Tests

College placement tests are used to decide which course is most suitable for students given their background or preparation. There are two classes of college placement tests. The first class, which we will refer to as remedial college placement tests

---

9 College admissions exams are not required of all incoming college students. Only examinees interested in attending a four-year postsecondary institution are required to take a college admissions test.
exams, is used to identify students who may require additional instruction in basic content. Because remedial college placement tests measure lower-level skills, items focus primarily on content from basic courses.

The second class of college placement exams is used more broadly to determine which courses students are prepared to take. This class of college placement tests is intended for all students enrolling in a given institution, from those who have taken no more than a basic course to students who have a strong course-taking background. Items on this test are designed to cover a wide range of proficiency levels, and assess content ranging from basic to advanced. Student placement into a course is usually determined by both the number and type of questions answered correctly.

College placement assessments vary by institutions within and across states. In our sample, many of the four-year postsecondary institutions, particularly the more selective ones, develop placement exams in accordance with their own department standards. In contrast, most two-year postsecondary institutions in our sample use commercially available assessments.

**End-of-Course Tests**

End-of-course tests assess student mastery of content from a particular course. For this study, they are considered a subsidiary of the above categories because they served the same purpose as the other types of tests. That is, in addition to measuring course-specific knowledge, the end-of-course tests in our sample were also used to monitor student achievement toward state standards, assess student readiness for postsecondary work, and place students in an appropriate college course. We have chosen to distinguish end-of-course exams as a separate category because their scope of content is much narrower than that of the other types of assessments.

**Expected Relations Among Tests with Different Functions**

Given that the test content should reflect its purpose (NRC, 1998), certain relations among the different types of assessments can be expected. Below we highlight discrepancies that can be predicted \textit{a priori} among the different types of exams in terms
of problem solving, inference, and advanced content items, and in terms of breadth of content coverage.

A Priori Discrepancies Among Different Types of Tests with Respect to Problem-Solving, Inference, and Advanced Content Items

Because they are used to discriminate among higher-achieving students, college admissions exams can be expected to contain the highest proportion of items assessing problem solving and inference. In contrast, remedial college placement tests, which are intended to assess entry-level skills, will contain the lowest proportion. Given the movement towards “tests worth teaching to,” state achievement exams are likely to include some problem-solving and inference items. However, because state achievement tests are also taken by examinees who are not interested in postsecondary education, these tests are unlikely to contain as many problem-solving and inference items as college admissions exams.

Advanced content items are likely to be most prevalent on college admissions exams and on college placement tests used to place students into a course commensurate with the student’s prior background. The presence of these items on college admissions exams is consistent with the tests’ purpose of sorting applicants well-qualified for college instruction from those less qualified. Likewise, because college placement exams are taken by students of a wide range of proficiency levels, including those examinees with a strong academic background, there needs to be advanced content in order to place well-prepared students into an appropriate higher-level course.

Advanced content, however, is likely to be devoid from remedial college placement tests, as the inclusion of such items would undermine the utility of remedial college placement exams as a measure of entry-level skills. Advanced content is also likely to be absent from state achievement tests that are used to award students a high school diploma because these tests need to maximally discriminate among the lower-regions of the achievement distribution.

The following discussion excludes end-of-course exams because the proportion of problem-solving, inference, and advanced content items depends upon whether the end-of-course exams are used to monitor student achievement, guide admissions decisions, or place students into an appropriate college course. In other words, the content of end-of-course exams depends upon the specific ways in which they are used, so we cannot make general comments about test content.
A Priori Discrepancies Among Different Types of Tests with Respect to Breadth of Content Coverage

Reflecting their function of assessing course-specific knowledge, end-of-course tests will have a narrow breadth of content coverage. Although remedial college placement exams will not have as narrow a focus as end-of-course tests, remedial college placement exams can also be expected to assess few topics, namely lower-level content areas. In contrast, state achievement tests, which are typically designed to provide a survey of student proficiency, will include many course topics. Likewise, college placement tests that are used to place students into a course commensurate with their prior background will contain a broad range of topics, from basic to advanced. Because readiness for postsecondary instruction requires knowledge across a range of content areas, college admissions exams will also have a wide breadth of content coverage.

Why is Test Alignment Important?

Tests can provide direction to students about the kinds of skills that are valued and should be developed (Bond & Roeber, 1998). However, if the competencies assessed on each exam (of the same subject matter) bear little relation to one another, students can receive confusing messages about the competencies that should be mastered in order to be prepared for college-level courses (Kirst, 1998; Education Trust, 1999). An example of this disconnect involves content differences between state achievement and college placement measures. Students may assume that passing a state achievement test and earning a high school diploma indicates a readiness for postsecondary instruction. However, because state achievement tests demand lower-level skills and knowledge than those required by college placement tests (Education Trust, 1999), students who perform well on a state achievement test may nevertheless lack the relevant content knowledge to perform well on college placement tests, thereby forcing them to enroll in remedial courses.\footnote{Discrepancies in performance between two assessments may be due to differences in content, cut scores, or both.} In an analogous manner, college admissions measures have been found to be less rigorous than college placement measures (Education Trust,
This means that the tests guiding admissions decisions is based on one notion of content deemed important for postsecondary instruction, but that the tests guiding course placement decisions is based on an entirely different notion (Kirst, 1999). Presumably, if the assessments administered to prospective college students were more aligned to one another, students would have a better sense of the kinds of skills that are necessary to be adequately prepared for college-level work.

Alignment has also been embraced as a means of reducing overall testing burden on students (Kirst, 2000). In light of estimates that students can spend upwards of two weeks solely on taking state achievement tests (McNally, 1999), both parents and policymakers have expressed concern over the possibility of a testing overload, especially for college-bound students (Committee for Economic Development, 2001; Firestone, 2000). Many states, including California, are exploring ways in which to streamline their assessment systems (see SB 233; California State Board of Education, 2001b; Manzo, 2001). Alignment may be one possible strategy because with greater consistency among assessments, various measures can be consolidated so that multiple uses can be derived from fewer tests. Specifically, if there were greater consistency between state achievement and college admissions tests, or between state achievement and college placement measures, then state achievement tests can be used in place of college admissions or college placement tests. The possibility that state achievement tests can inform college admissions or postsecondary course placement decisions will be explored further in the case study reports.

The Case Against Test Alignment

While alignment among different assessments may be desirable in some instances, it is not always possible, nor desirable. As explained earlier, differences in test content and format may reflect variations in test function. Furthermore, even when exams are designed to be parallel, as in alternate forms of the SAT I, we would not

---

12 On average, college placement tests were found to assess intermediate algebra and trigonometry to a greater extent than did college admissions exams.

13 This is not to imply that different entities make decisions about admissions and postsecondary course placement decisions. The same college makes both decisions.
expect perfect alignment. Because tests are merely samples from a larger universe of possible items, they are likely to differ from one to the next.

Another reason why greater alignment among tests has not been universally espoused is because tests that differ in content are needed to serve as auditing mechanisms to corroborate score increases on high-stakes measures. As tests assume higher stakes, there will be increased efforts to raise scores. Some of these efforts may be inappropriate and lead to artificial score gains (Shepard & Dougherty, 1991). Concerns about inflated score increases has prompted educators and policymakers to call for a confirmation of progress via concordance of score trends on other assessments (see e.g., President George W. Bush’s “No Child Left Behind,” 2001; Linn, 2001). These other assessments should measure somewhat similar skills and knowledge as the high-stakes tests in order to determine the generalizability of the gains (i.e., if the gains represent substantive increases in proficiency, the gains should be reflected to a substantial degree on the other assessments). However, the other assessments must also contain sufficiently different content so as to be less susceptible to the factors that lead to score inflation on the high-stakes exams. Thus, some content differences among tests need to be maintained so as to verify that score gains observed on high-stakes tests are not spurious.

**Defining Misalignment**

As noted above, aligned measures may hold some benefits, but perfect alignment among tests is neither likely nor desirable. In this study, we use the term “misalignment” to refer only to large discrepancies that cannot be attributed to differences in test use, and therefore send students mixed messages about the kinds of skills that are valued. Thus, discrepancies that can be predicted *a priori* are not misalignments.

To illustrate an example of a misalignment, consider two hypothetical college admissions tests that are used to inform admissions decisions. Suppose that both tests require students to produce a writing sample, but that the scoring rubrics differ so that one test penalizes students for punctuation or spelling errors, whereas the other test does not. This difference between the tests’ scoring criteria would be considered a misalignment, as the discrepancy is not due to variations in test purpose.

---

14 In Chapter 2, we will define the standards for judging “large” discrepancies.