Summary

Remediation at the postsecondary level has attracted considerable attention in recent years. The nearly 30% remediation rate among first-time college students nationwide has prompted concern among many policymakers and educators because students requiring remediation are less likely to earn a degree, and estimates of the annual budgetary costs of remedial programs to the nation’s postsecondary institutions have reached as high as one billion dollars.

Educational policymakers have suggested the high level of remediation at the postsecondary level is partly due to differences in the content, format, and scope of assessments administered to prospective college students. Assessments provide direction about the competencies that should be developed, but because tests (of the same subject matter) administered to prospective college students do not necessarily measure the same kinds of skills and knowledge, students may receive conflicting signals regarding the competencies needed to be prepared for college-level work. In an effort to better articulate the competencies that students should master in order to enroll in non-remedial college courses, policymakers have recommended increasing the alignment among different measures.

Alignment has also been embraced as a means of reducing overall testing burden on students. In light of concerns by policymakers and parents that college-bound students are overtested, many states are exploring ways in which to streamline their assessment systems. 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.

Perfect alignment among tests, however, may not be possible. Because test content is indicative of the intended use of a measure, tests that have different purposes will vary somewhat. Thus, differences in test use necessarily limit the extent to which tests can be aligned.

The primary purpose of this study is to describe alignment among different assessments administered to secondary and incoming first-year college students in five
case study sites: California, Georgia, Maryland, Oregon, and Texas. The analysis also examines the implications of discrepancies among exams, and explores the possibility that a test can be used for other purposes beyond the uses for which it was specifically designed.

Research Design

*Purposes of Assessments Examined in this Study*

We restricted our analysis to math and English/Language Arts (ELA) measures administered to high-school or incoming first-year college students. No assessments administered at the postsecondary level were examined. The assessments were limited to those used by selected institutions in our five case study sites. We chose institutions to represent a range from less selective to highly selective. The selection, however, is not a scientific sample. More information about the chosen institutions is provided in the case study reports (Chapters 3-7).

We examined four types of assessments in this study: state achievement, college admissions, college placement, and end-of-course exams. State achievement tests provide a broad survey of student proficiency toward state standards in a particular subject like math or reading. College admissions exams sort applicants more qualified for college-level work from those less qualified. College placement tests determine which course is most appropriate for students. Some college placement tests are used to identify students who may require additional remediation, whereas other college placement exams are used more broadly to determine which course students are eligible to enroll in, given their prior academic preparation. End-of-course tests are measures of knowledge of one particular course.

*Defining Misalignment*

In this study, we use the term “misalignment” to refer 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.

---

1 The case study sites were chosen for a variety of state-specific reasons. Readers interested in more details of the case study selection factors are referred to Chapter 2.
Methodology

Two raters judged alignment among assessments using three coding categories. The coding categories describe the technical features, content, and cognitive demands of each assessment.

Math Coding Categories

Technical features involve characteristics such as time limit and format. For the content category, raters coded each item as assessing one of several math content areas, ranging from basic through advanced math. Content areas include prealgebra, elementary algebra, intermediate algebra, planar geometry, coordinate geometry, statistics and probability, trigonometry, and miscellaneous. To evaluate the cognitive demands of a test, raters coded each item as measuring conceptual understanding, procedural knowledge, or problem solving.

ELA Coding Categories

Coding categories in ELA cover three types of skills: reading, editing, and writing. Reading skills relate to students’ vocabulary and comprehension of reading passages. Editing skills relate to students’ ability to recognize sentences that violate standard written English conventions. Writing skills pertain to how well students can produce a composition that clearly and logically expresses their ideas.

Technical features describe characteristics such as time limit and format. The content category has three dimensions and describes the subject matter of the reading passage (topic), the author’s writing style (voice), and the type of reading passage (genre). Topic consists of five levels—fiction, humanities, natural science, social science, and personal accounts. Voice consists of four levels—narrative, descriptive, persuasive, and informative. Genre also contains four levels—letter, essay, poem, and story.

For exams measuring reading or editing skills, raters used all three dimensions of the content category to describe the reading passages. However, all three dimensions are not relevant to writing tests, as such tests do not include reading passages. Instead,
writing measures contain a short prompt that introduces a topic that serves as the focus of students’ compositions. Only the topic dimension is used to describe the writing prompts.

For measures of reading and editing skills, the cognitive demands category distinguishes among three levels of cognitive processes: recall, evaluate style, and inference. For writing tests, the cognitive demands category focuses on the scoring criteria, particularly the emphasis given to mechanics, word choice, organization, style, and insight.

Results

Below, we highlight the general patterns of results across the five case study sites. Readers who are interested in specific details of a state are referred to the case study chapters (Chapters 3 – 7).

Math

All tests, regardless of purpose, use multiple-choice items to assess student knowledge. College admissions exams, as well as college placement tests used to place students into a course commensurate with the students’ prior math background, assess advanced content (i.e., intermediate algebra and trigonometry) to the greatest extent. In contrast, remedial college placement tests contain the largest proportion of items assessing basic math skills (i.e., prealgebra). College admissions and state achievement tests sample broadly across several areas of math, but end-of-course exams are focused on a single content area, either algebra or geometry. With respect to cognitive demands, college admissions exams contain, on average, the highest proportion of problem-solving and conceptual understanding items, whereas remedial college placement tests contain the fewest.

---

2 Evaluate style relates to improving the organization, clarity, or presentation of the item or passage.
**ELA**

**Reading**

The majority of reading measures assess reading proficiency solely with multiple-choice items. Regardless of test purpose, reading passages are typically narrative or informative essays. Although topics vary greatly from one test to the next, humanities, fiction, and personal accounts are the more popular topics. With respect to cognitive demands, college admissions tests are, on average, more likely than either college placement or state achievement tests to assess inference skills.

**Editing**

Few discrepancies are observed among editing measures. All editing measures assess students’ knowledge of standard written English solely with multiple-choice items. Reading passages on editing measures are typically narrative or informative essays on a humanities topic. In terms of cognitive processes, tests of all purposes tend to assess evaluate style and/or recall skills.

**Writing**

Few college admissions exams or college placement tests require students to produce a writing sample. In contrast, the majority of state achievement tests requires a composition. Virtually all writing tests include mechanics, word choice, style, organization, and insight as part of their scoring criteria, but some measures (most notably the SAT II Writing) exclude insight from their scoring rubrics.

**Discussion**

*Is There Evidence of Misalignment?*

In math, there are no instances of misalignments, as discrepancies among assessments reflect variations in test use. Because remedial college math placement tests are used to identify students who need additional development in basic math skills, test content is focused on lower-level areas such as prealgebra. Items assessing problem-solving and advanced content are most common on college admissions exams because
these types of items help distinguish examinees more qualified for college-level work from those less qualified. College placement exams that are used to determine which course is most appropriate for students given their prior preparation must accommodate students of a wide range of proficiency levels, including those with strong math backgrounds. Therefore, advanced content such as intermediate algebra and/or trigonometry are included so that well-prepared students can be placed into higher-level math courses. Consistent with their purpose as a broad survey of student proficiency, most state achievement measures assess an array of content areas. End-of-course exams, on the other hand, are designed to assess proficiency of a particular course, and therefore narrow their content to primarily a single area.

As in math, the majority of discrepancies among ELA assessments are not misalignments because they reflect differences in test use. However, one notable misalignment pertains to the scoring criteria of the writing measures. Most writing exams include insight as part of its scoring rubrics, but a few measures do not. Ideally, if the developers of these exams were to add insight to the scoring criteria, or provided a clear rationale of why insight has been omitted from the scoring rubrics, students would receive a more consistent signal about the importance of insight with respect to writing proficiency.

Can State Achievement Tests Inform Admissions or Course Placement Decisions?

We examined the possibility that state achievement tests can be used for purposes other than monitoring student achievement toward state standards. Generally, state achievement tests can be used as alternatives to remedial college placement exams. However, most state achievement tests cannot inform college admissions or broad course placement decisions (i.e., deciding which course is most appropriate for students, given the students’ prior preparation) because they contain, on average, fewer problem-solving, inference, or advanced content items than do college admissions and college placement tests.
Conclusion

Although discrepancies among assessments in both math and ELA are numerous, differences generally reflect variations in test purpose. Because test purpose dictates the kinds of content that is included, attempting to align tests without regards to test function may seriously undermine the usefulness of those measures.