

# College Admission Tests as Measures of High School Accountability

Richard J. Noeth, David J. Rutkowski, and Bridget A. Schleich

AUGUST 2010

## CONTENTS

**Introduction..... 1**

**Learning Standards and Assessments ..... 1**

**College Admission Tests as Statewide Accountability Measures..... 2**

**Admission Test Issues for State Policymakers..... 3**

**Illustrations of Admission Test Use at the Statewide Level..... 4**

**Summary and Recommendations..... 9**

**Authors..... 10**

**Acknowledgements ..... 10**

**References..... 10**

**Web Resources ..... 12**

## INTRODUCTION

The use of national college admission tests to assess the achievement of state high school learning standards and high school accountability has become a growing phenomenon in American education. Yet, although this test application has been adopted by a number of states, there has been little independent analysis and evaluation of its appropriateness and validity for all constituents.

This CEEP Special Report is intended to frame a series of key baseline issues (and offer corresponding recommendations) for state policymakers, practitioners, and researchers as they consider and evaluate the current and/or future use of traditional admission tests to assess achievement and accountability. Specifically, the report addresses the following question from multiple perspectives:

*Should tests that were originally intended and developed to serve nationwide college admission and placement purposes, and were designed for national college-going populations, now also be used to assess the achievement of state learning standards and high school accountability?*

This report presents a series of interrelated sections that focus on the background, use, and primary issues related to the statewide application of college admission tests. Included are:

- A brief review of high school learning standards and the measures developed to assess the achievement of these standards.
- Multiple perspectives on the use of admission tests as sole or partial measures of high school achievement and accountability.

- A set of specific issues that focus on the application of admission tests to assess student achievement and high school accountability.
- Illustrations of Admission Test Use at the Statewide Level that outline specific uses of admission tests to assess state learning standards (the Illinois Prairie State Achievement Examination [PSAE] which uses the ACT, and the Maine High School Assessment [MHSA] which uses the SAT Reasoning Test).
- Recommendations for state policymakers, educators, and researchers as they consider and evaluate the present and/or future use of admission tests to assess achievement and accountability.

## LEARNING STANDARDS AND ASSESSMENTS

Educational standards are a dominant feature of American education, and are often the central framework guiding state education policy (Shepard, Hannaway, & Baker, 2009). The No Child Left Behind (NCLB) Act requires that all states have in place educational achievement and academic content standards in language arts, mathematics, and science at the elementary, middle, and high school level (U.S. Department of Education, 2003). While states have largely developed their own standards to comply with this Act, the federal government has provided guidelines that standards should be conceptualized as a system to include achievement levels, achievement descriptors, exemplars, and achievement-level cut scores (U.S. Department of Education, 2003).

NCLB also mandated that states create standards-aligned assessment programs to ensure that standards are being met. An aligned assessment must evaluate state

learning standards and ensure that the following dimensions are met: comprehensiveness, content and performance match, emphasis, depth, consistency with achievement standards, and clarity for users (U.S. Department of Education, 2003). The alignment to standards is important so that assessment guides instruction in ways that are intended by the standards (Linn, 2005), and because assessment has a stronger impact on what is taught in the classroom than mandated standards alone (Hamilton et al., 2007).

---

***...the use of norm-referenced tests that have not been originally designed to assess specific standards has become a practice in a few states as a means to gauge NCLB compliance.***

---

In addition to alignment, the interpretive framework of an assessment is also important. Interpretive frameworks within state NCLB-compliant assessments generally fall into two categories: criterion-referenced (designed to provide a measure of performance that is interpretable in terms of a clearly defined and delimited domain of learning tasks) and norm-referenced (designed to provide a measure of performance that is interpretable in terms of an individual's relative standing in some known group) (Linn & Miller, 2005).

Assessments can also be classified by the level of stakes placed on them. Stakes are often classified as:

- High stakes - which have direct consequences attached to results; for example, assessments used to determine graduation, promotion, college admission, or a scholarship (Payne, 1997), or those assessing a school's Adequate Yearly Progress (AYP).
- Low stakes - which have little to no consequences outside the school, but possibly involve classroom consequences (Mitchell, 2006).

Each type of assessment can incorporate a number of measures across content areas. Multiple measures can support a range of opportunities (retesting, longitudinal testing), formats (multiple choice, constructed response, true-false, essays), domains (different spheres of student achievement), and sources (a combination of internal and external test sources) (Shafer, 2003).

Although measures and standards often go hand in hand, their development and application can vary. Most educational standards tend to be developed to express what the authoring group feels should be valued. Once these standards are in place, measures are then determined and applied to each standard. Although a large number of measures are developed after standards are in place, standards can also be matched to an already-existing measure.

The majority of states base their assessments on criterion-referenced tests designed to determine their respective standards. However, the use of norm-referenced tests that have not been originally designed to assess specific standards has become a practice in a few states as a means to gauge NCLB compliance. For example, three states have incorporated national college admission tests (classified as norm-based assessments) as part of their NCLB assessments (Illinois, Maine, and Michigan) and at least three other states incorporate them into low-stakes accountability systems (Colorado, Kentucky, and Wyoming).

---

### **COLLEGE ADMISSION TESTS AS STATEWIDE ACCOUNTABILITY MEASURES**

A fundamental challenge to the use of admission tests to evaluate a student's coursework achievement or a secondary school's performance was issued by the Commission on the Use of Standardized Tests in Undergraduate Admission<sup>1</sup> (National Association for College Admission Counseling, 2008). Labeling this practice as "test misuse," the Commission recommended that states refrain from using unmodified admission tests, especially when the tests are used for high-

---

1. Hereafter referred to as the Commission.

stakes accountability purposes. The Commission stated that admission tests were not designed for this purpose, and are not sufficiently tailored to assess progress toward explicit measures for learning in a given state.

Citing the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education, 1999), the Commission indicated that admission tests, whether they are intended to assess achievement or ability, are not directly linked to a particular instructional curriculum and are not appropriate for measuring high school performance.

Many individual standards in the Standards for Educational and Psychological Testing, as well as individual principles that focus on appropriate test use in the Code of Fair Testing Practices in Education (Joint Committee on Testing Practices, 2004), are directly applicable to the use of admission tests as statewide achievement and accountability measures. Related key individual standards (abridged) include:

- A rationale should be presented for each recommended interpretation and use of test scores, together with a comprehensive summary of the evidence and theory bearing on the intended use or interpretation (Standard 1.1).
- When unintended consequences result from test use, an attempt should be made to investigate whether such consequences arise from the test's sensitivity to characteristics other than those it is intended to assess or to its failure to fully represent the intended construct (Standard 1.24).
- If local examinee groups differ materially from the population to which norms refer, users who report derived scores based on the published norms have the responsibility to describe such differences (Standard 4.7).
- Those who mandate the use of tests should monitor their impact, and identify and minimize potential negative consequences; consequences resulting from the uses of the test (both intended and unintended) should be examined by the test user (Standard 13.1).
- When a test is used with respect to specified curriculum standards, evidence of the extent to which the test samples the range of knowledge and elicits the processes

reflected in the target domain should be provided—including those aspects it fails to represent (Standard 13.3).

Directly related to the appropriateness of using admission tests to assess achievement and accountability is the issue of alignment. Specifically, what is the degree of alignment between an admission test, developed and normed for a national college-going population, and a state's set of learning standards for all students? How is this alignment evaluated and monitored as tests and standards evolve over time? Achieve and the Education Trust (2008) reported that college admission tests are not designed to be given as high school tests with multiple purposes, such as determining if all students are ready for credit-bearing coursework or holding schools accountable for meeting college and career readiness standards, largely because admission tests are not designed to assess the curriculum of any particular school or state.

---

***... admission tests are not designed to assess the curriculum of any particular school or state.***

---

The Southern Regional Education Board (2007) indicated that a state should use college admission tests to determine academic readiness only if the standards on which they are based are incorporated fully into the state's high school academic standards. Further, states should consider how secondary school standards and assessments are aligned with state postsecondary readiness standards. Research suggests that states' learning standards and related assessments, particularly as related to college readiness, are often out of alignment and that significant gaps may exist between what states expect to be assessed and what is being assessed (e.g., Brown & Niemi, 2007; Callan, Finney, Kirst, Usdan, & Venezia, 2006).

The Commission was influenced by an Achieve (2007) report that examined college admission tests and recommended that

states augment these measures when incorporating them into statewide testing systems. Achieve reported that there were gaps in what admission tests assess, and indicated that neither the ACT nor SAT included the full range of advanced concepts and skills reflected in the American Diploma Project benchmarks, and in state high school standards. Conley (2007) suggested that although admission tests have functioned as methods of identifying students who are potentially college-ready, advances in the understanding of the key knowledge necessary to succeed in college courses suggest a potentially different, or at least supplementary, measure of content knowledge.

Achieve and the Education Trust (2008) discussed the use of modified admission tests as a viable state assessment approach, and described several important challenges inherent in this approach:

- There may be unclear alignment to state standards.
- There may be development and administration costs to augment tests.
- The full range of advanced concepts and skills is not always included.
- Testing time in the school day may be increased.
- There may not be enough feedback to students and schools about college readiness to inform grade 12 coursework.<sup>2</sup>

In light of the above, Achieve and the Education Trust suggested that states choosing to modify admission tests should work with ACT and the College Board to adapt these tests, encouraging ease of administration and greater coherence and alignment with state standards.

---

### **ADMISSION TEST ISSUES FOR STATE POLICYMAKERS**

This section focuses on a series of critical issues to be considered and weighed in determining whether a national college admission test can effectively serve as a valid and appropriate measure of student achievement of state high school learning standards and school accountability.

2. The PSAE and MHSA are administered to grade 11 students in the spring. An overview of each assessment is provided in the Illustrations of Admission Test Use at the Statewide Level (pp. 4-5).

---

## **Goals and Evaluation of Statewide Admission Test Use**

A state contemplating or evaluating the use of an admission test to assess student achievement and school accountability should have a specific and documented set of initial and ongoing goals directly related to this measurement application. These goals should represent a broad range of constituent input, including policymakers, administrators, business leaders, teachers, learners, and parents. Additionally, there should be in place an initial and ongoing evaluation plan to formally determine whether using an admission test is meeting established and evolving goals.

Achieve and the Education Trust (2008) reported several distinct advantages of using college admission tests to assess achievement of state high school learning standards. These included widespread public acceptance and recognition, use in college admission and placement decisions, national normative comparisons, national portability, and encouragement and support in the college preparation process.

ACT and the Illinois State Board of Education (2006) established a set of goals for use of the ACT (and WorkKeys) and set eight necessary conditions to determine whether these measures could be appropriately used as part of the PSAE. Included were:

- The ACT and WorkKeys must assess the Illinois standards.
- The use of the ACT and WorkKeys should be consistent with the intended outcomes of the Illinois assessment programs.
- Neither the ACT nor WorkKeys should be used by themselves as the sole criterion in making high-stakes decisions about students, school effectiveness, or teacher effectiveness.

The Maine Department of Education indicated that the transition from the Maine Educational Assessment to the SAT Reasoning Test was based on the need to move the public education system to a higher level of effectiveness (Gendron, 2006). Among the policy goals supporting this change were:

- To build an assessment program that increases student motivation to do well on the test and to think of the assessment pro-

gram as an important tool in preparation and planning for college.

- To ensure that the assessment program is carefully coordinated with other initiatives undertaken by state government and partnering organizations designed to increase the number of students that go on to postsecondary education.
- To strengthen curriculum and instruction to minimize the extent to which students require remedial coursework upon entering college.

In terms of formal goal evaluation on the use of admission tests at the statewide level, ACT (2009a) created case studies for Illinois and Colorado that focused on improvements in academic achievement, readiness, number of students considering college, and college enrollment and retention. However, there appears to be little information on state criteria set for goal evaluation, as well as actual formal goal evaluation done by the states that have adopted admission tests (either subsequent to their initial administration or on an ongoing basis).

## Alignment with State Learning Standards

A critical issue related to the use of admission tests to evaluate student performance and school accountability centers on standards-assessment alignment. Although states that have implemented this test application have passed federal alignment reviews, there remain concerns among reviewers that not all standards are being appropriately assessed.<sup>3</sup>

For example, in Illinois a number of initial alignment studies were completed for the PSAE (ACT & the Illinois State Board of Education, 2006). These studies found acceptable alignment between the Illinois learning standards and ACT/WorkKeys, and called for improved alignment between the science assessment and science standards. There also were concerns surrounding the mathematics component where a number of recommendations and observations found that:

3. Additional information on PSAE and MHSA alignment studies can be found in the Illustrations of Admission Test Use at the Statewide Level (pp. 4-5).

## Illustrations of Admission Test Use at the Statewide Level

The following illustrations, featuring the Illinois Prairie State Achievement Examination (PSAE) and the Maine High School Assessment (MHSA), outline each state's specific use of admission tests to assess learning standards and school accountability.<sup>a</sup>

**PSAE.** The PSAE is aligned to Illinois high school learning standards in reading, mathematics, science, and writing to assess progress toward meeting the standards for state and federal accountability requirements and to recognize the achievement of students who earn a Prairie State Achievement Award. Although graduation is contingent upon taking the PSAE, students are not required to pass the examination in order to receive a diploma. The PSAE is comprised of the ACT Plus Writing (the English, mathematics, reading, science, and writing tests) and portions of WorkKeys (the applied mathematics and reading for information tests). Augmenting these assessments is the Illinois State Board of Education (ISBE) designed science assessment.

Several sequential alignment studies were completed to ensure that the PSAE met the NCLB standards-assessment alignment requirements. The following comprise a brief description of these studies:

1. An ACT study reported that five of the six reading and writing Illinois learning standards and 16 of the 18 standards in mathematics were covered. Science standards aligned with the ACT and the ACT College Readiness Standards were also subsumed under the Illinois learning standards.
2. ISBE studies reported sufficient alignment between Illinois learning standards and the ACT and WorkKeys in mathematics and language arts; and although there were gaps in the ACT Science assessment, this was ameliorated by the addition of the ISBE-developed science assessment.
3. A third set of studies generally concluded that although adequate alignment existed between the PSAE and the learning standards, there were some discrepancies including a mismatch in English language arts and gaps in mathematics.
4. An independent study assessed alignment in reading, mathematics, and science, and found that in every case, discrepancies were acceptable.

The PSAE is administered toward the end of the junior year and covers subjects assessed by three separate examinations spread over two consecutive days. On the first day, the ACT Plus Writing assessment is administered. The second day consists of the ISBE-developed science assessment and WorkKeys assessments.

PSAE scores are scaled and proficiency cut points specify: Academic Warning, Below Standards, Meets Standards, and Exceeds Standards. Students also receive an ACT score that can be used for college admission and placement purposes as well as a set of WorkKeys scores that can be provided to future employers. For the 2009 assessment, 53 percent of participants met or exceeded Illinois education standards. (An identical proportion of students met or exceeded standards in 2007 and 2008.) Illinois did not achieve the NCLB-mandated Adequate Yearly Progress (AYP) in 2009 and has failed to show improvement in students who meet or exceed required scores on the PSAE since 2002.

<sup>a</sup> All information included in both the PSAE and MHSA illustrations is directly taken from the resources provided on each assessment's website. These respective websites are <http://isbe.state.il.us/assessment/psae.htm> and <http://www.maine.gov/education/mhsa/index.htm>



Given mandates to meet NCLB Adequate Yearly Progress, the PSAE functions as a high-stakes assessment for schools, districts, and the state. NCLB holds schools, districts, and states accountable insofar as monetary funding and avoiding other sanctions are contingent on achieving AYP. The stakes for students are less clear in that failing the PSAE does not prohibit graduation; however a student's final transcript includes PSAE scores. Additionally, the PSAE is linked to a number of scholarships and state awards.

**MHSA.** Maine requires all juniors to take the writing, critical reading, and mathematics portions of the SAT Reasoning Test to assess progress under NCLB. To address alignment concerns, the College Board conducted a study between Maine learning results (i.e., standards) in English language arts and mathematics and the SAT. The study concluded that in many cases, the SAT aligned well with learning results; however, the College Board recommended the augmenting of test questions in both content areas to achieve satisfactory alignment with the results. A subsequent study resulted in mixed findings, with misalignment in a number of areas. Citing these two alignment studies, the U.S. Department of Education initially did not approve Maine's request to use the SAT as their sole source for grade 11 NCLB assessments.

Maine augmented the SAT with a mathematics assessment and a science assessment, and subsequent satisfactory alignment studies resulted in final approval from the U.S. Department of Education. Maine recently adopted the New England Common Assessment Program (NECAP) standards, and given this adoption and new alignment studies, determined that a mathematics augmentation was no longer required be in compliance with NCLB requirements.

The MHSA is administered near the end of the academic year over two non-consecutive days. The science portion is administered during the school day within a two-week window and the SAT is taken on the first Saturday in May. All students receive two sets of scores: first a traditional set of SAT scores and following that a set of MHSA scores based on a weighted combination of the science test and the SAT.

The MHSA reporting categories are: Does Not Meet Standards, Partially Meets Standards, Meets Standards, and Exceeds Standards. To assess AYP under NCLB, school and higher level reporting is based on the proportion of students who meet or exceed standards. The Maine Department of Education reported that for 2008, 46 percent met or exceeded standards in reading, and 40 percent met or exceeded standards in mathematics. Compared to a target of 64 percent in reading and 43 percent in mathematics, Maine was short of its goal.

The stakes associated with the MHSA are dependent on the particular stakeholder. Student-level stakes are generally considered low because graduation is not contingent on test performance, nor are scores reported on transcripts. Stakes are higher for schools, districts, and the state, as funding can be reduced and in some cases school management can be subject to restricting or removal based on less than adequate AYP.

---

*For a detailed policy analysis of the use of the ACT as a component of the PSAE, please go to:*

[http://ceep.indiana.edu/projects/PDF/Special\\_Report\\_ACT\\_Policy\\_2010.pdf](http://ceep.indiana.edu/projects/PDF/Special_Report_ACT_Policy_2010.pdf)

- The PSAE should consider a better balance across content areas by including number and operation basic skills in problems involving measurement, geometry, and algebra.
- The PSAE does not fully assess a student's ability to support answers through reasoning and evidence.
- Use of multiple-choice format does not provide students the opportunity to formulate their own responses and communicate their findings in writing (Dossey & McCrone, 2006).

Regarding the reading and writing portions, reviewers found that the PSAE possessed strong alignment; but not all of the Illinois Learning Standards for English language arts are addressed by the PSAE, nor can they be appropriately addressed in a two-day, timed, paper-and-pencil examination (Ogle & Hunter, 2006).

Similarly, Maine's initial alignment studies for language arts and mathematics indicated that the assessment did not adequately evaluate standards and consequently did not meet federal requirements (Gendron, 2006). These studies resulted in mixed findings with misalignment in a number of areas (College Board, 2005; Webb, 2006). Maine then augmented the mathematics portion of the SAT and further evaluations found that the alignment situation improved (Maine will not include mathematics augmentation for the 2010 assessment after adopting a new set of standards).

Alignment issues highlight a number of related factors for states to consider: misalignment between the assessment and learning standards, too few questions of a desired format, and an over- or under-concentration of questions in a particular area. Despite federal guidelines to assist state policymakers, there remains the concern that any test not specifically designed for the purpose of assessing learning standards may not be the most appropriate choice for an accountability system. Although alignment studies assist in ameliorating these concerns, they also indicate where improvements and adjustments are necessary.

---

## Evolving State Standards and Admission Test Content

Directly related to the issue of alignment between the content of an admission test and state learning standards is the implication of changes in one or both over time; and consequently how alignment is affected. Even if there is a documented initial alignment between what an admission test assesses and the standards established in a state, there exists the likelihood that one and/or the other of these elements will evolve over time, for reasons essential to their primary purposes and constituencies.

More likely, this evolution will take place within state standards with changes in subject matter, pedagogy, technology applications for learning, and postsecondary and workforce requirements — and the state's response to these exigencies. Far less likely, yet historically documented, is the ongoing evolution of admission tests — in terms of breadth and depth of content, format, difficulty level, measurement applications, and responsiveness to changing national postsecondary education (and possibly workforce) demands.

For example, over the past decade several curriculum surveys have been undertaken by ACT to guide the development and modification of the ACT (and ACT's two other Education Planning and Assessment System's measures, EXPLORE and PLAN). The conclusions of these surveys have been that:

- No changes in either the constructs or the test specifications are warranted (ACT, 2003).
- The tests reflect current instructional practices and college readiness expectations (ACT, 2006a).
- The tests appropriately reflect college readiness expectations across the areas of English/writing, mathematics, reading, and science (ACT, 2009b).

Similarly, the most recent changes to the SAT were made in the spring of 2005 with the implementation of the SAT Reasoning Test, designed to enhance test alignment with current high school curricula and emphasize the skills needed for success in college (Mattern, Patterson, Shaw, Kobrin, & Barbuti, 2008). The previous changes to the SAT came a decade earlier with the

introduction of the SAT I (Lawrence, Rigol, Van Essen, & Jackson, 2002).

As a whole, Illinois Learning Standards have not changed; however, new supplementary materials, performance descriptors, and assessment frameworks have been developed to help clarify the standards. Conversely, the Maine Department of Education has completely redeveloped their educational standards by recently adopting standards set forth by the New England Common Assessment Program (NECAP).

Thus a key consideration is not only the need to initially determine the degree of alignment between admission test content and state learning standards, but also to evaluate what implications possible standards and assessment changes have for ongoing alignment. Admission tests have a history of infrequent, but important, changes over time to reflect evolving national postsecondary demands, based upon changing postsecondary readiness standards and requirements. The criteria for evolving state standards and state responsiveness to these changing needs have been far more dynamic.

---

## Sole or Partial Admission Test Use and Component Weighting

When integrating admission tests into an accountability framework, the issue of using part or all of the assessment emerges. Both Maine and Illinois have either fully or partially met the American Diploma Project's (Achieve, 2007) call for the augmentation of admission tests used for statewide testing systems. Initially in both states, augmentation ensured that federal guidelines of proper alignment were met. However, the MHSA recently removed the mathematics augmentation leaving the SAT as the singular assessment for language arts and mathematics.

Each of the PSAE subjects is covered by two separate assessments. The ACT is augmented in mathematics and language arts by WorkKeys, and science is augmented by an Illinois-designed assessment. Thus, for each content area, Illinois must integrate questions from two different assessments. Initially, the PSAE weighted

WorkKeys and the ACT so that each assessment equally contributed to a student's score. Following a recommendation from the U.S. Department of Education, the PSAE now uses un-weighted raw scores so that each question, regardless of the assessment to which it belongs, contributes equally to a score (Illinois State Board of Education, 2008). Illinois' adjustment from weighted to un-weighted scores is appropriate since each question, rather than the assessment in its entirety, is aligned to a learning standard.

However, several issues arise regarding the PSAE approach. Namely, what factors should be considered when one test is used to augment another (i.e., do the ACT and WorkKeys assess sufficiently different constructs and learning standards, and are customized, state-developed questions a more reasonable choice for augmentation)? For example, ACT (2006b) reported that preparation necessary for both college and workforce training programs is comparable, as are the related components of the ACT and WorkKeys, which raises the question of whether these two tests (at least in their entirety) are actually necessary.

---

***Concerns related to using part or all of a pre-existing assessment, augmenting assessments, and decisions associated with weighting assessments are important considerations for policymakers that may carry significant consequences.***

---

A number of policy issues emerge once a state has decided to augment an existing assessment. For instance, does using only specific parts of an admission test to assure shorter testing times outweigh the advantages of administering an entire test that results in usable ACT or SAT scores? Also, when a pre-existing test is not sufficient to cover all of a state's standards, augmentation by other assessments (using complete tests or a number of test questions) is often

recommended. An important decision when multiple assessments are used is how to weight each assessment (e.g., the choice between weighting the test questions or the entire assessment opens an opportunity to emphasize specific content areas or certain questions). And with assessment weighting comes the issue of how the primary assessment and the augmenting assessment will be considered when generating final scores (i.e., related to the relative weights of test components or individual questions). Concerns related to using part or all of a pre-existing assessment, augmenting assessments, and decisions associated with weighting assessments are important considerations for policymakers that may carry significant consequences.

---

## Impact on Underserved Populations

Underserved populations have faced many obstacles to realizing a successful educational experience, particularly in transitioning to the postsecondary level. One issue has been performance on college admission tests; results have consistently varied across student populations, with underserved students achieving well below the national average.

For 2009 college-bound test takers, 28 percent of Caucasian and 36 percent of Asian-American college-bound seniors met all of ACT's College Readiness Benchmarks. In contrast, only 4 percent of African-American and 10 percent of Hispanic college-bound seniors met these benchmarks (ACT, 2009c). The SAT reports similar differential results (College Board, 2009).

Yet, despite differences in performance across college-going populations, research indicates that both the ACT and SAT assess performance with little or no bias<sup>4</sup> regarding race, ethnic background, gender, and social class (National Association for Col-

---

4. The American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education (1999) described bias as follows: In a statistical context, a systematic error in a test score. In discussing test fairness, bias may refer to construct underrepresentation or construct-irrelevant components of test scores that differentially affect the performance of different groups of test takers.

lege Admission Counseling, 2008). Commensurately, there is a substantial body of research that has investigated and specified a number of reasons for these score disparities (Noeth, 2009; Zwick, 2007).

Thus, although admission tests have demonstrated technical fairness and there are plausible and documented reasons for differential results, the consistently poor performance of diverse populations raises an important issue related to admission test use as measures of all-student achievement and school accountability. Specifically, although there is information regarding differential performance of traditional college-bound student groups on admission tests, there has been little evaluation, research, and documentation regarding reasons for the differential performance of various all-student populations on these same measures. From this specific perspective, there might be insufficient information regarding the performance, and the factors related to that performance, for diverse all-student populations on admission tests as used in statewide assessments.

---

## Reconciling Potential Unintended Outcomes

There is a likelihood of unintended outcomes when using any assessment — whether such results encompass factors that impact performance, test interpretation, and/or the application of test information. This potential may be more likely when a test designed for one purpose and one population is then also used for additional purposes and other populations.

### The Implications of Test Preparation.

With regard to admission tests, there is a longstanding issue of the effects of special test preparation on test performance. When an admission test is used for accountability, there is also concern regarding how formal test preparation activities (possibly on a large scale) impact test performance. Further, it is important to consider how assessment preparation programs ameliorate or exacerbate (differential) levels of student achievement.

Formal test preparation or coaching is typically provided by a teacher, mentor, or program instructor who places emphasis on the teaching of specific test-taking strategies

and/or in-depth coverage of test content. Test preparation and coaching often involves a considerable amount of time and programs range dramatically in cost (Noeth, 2009). There is considerable evidence (Briggs, 2009; Zwick, 2007) that test preparation programs raise test scores; and these increases tend to be in the magnitude of 0.5 - 1.5 points on various ACT subtests and/or the ACT Composite score, and 20-30 points on the combined SAT components.

About 70 percent of high schools offer test preparation resources of some type. Further, a substantial percentage of high school students attend commercial test coaching programs, and these students typically come from families with more formal education and higher incomes (Noeth, 2009). However, this still leaves a large percentage of students without access to formal test preparation — perhaps due to schools not having the resources to offer such programs, high costs of commercial programs, or lack of time because of the need to work.

---

***Thus an important issue for state policymakers is to determine the degree of differential benefit that formal test preparation programs might offer various segments of the state's high school population ...***

---

Thus an important issue for state policymakers is to determine the degree of differential benefit that formal test preparation programs might offer various segments of the state's high school population (e.g., students from high-income v. low-income families, or students from schools with substantial resources v. schools with limited resources); and how that differential access might relate to differential achievement.

Additionally, the implications for differential access to preparation programs should be examined in terms of the evaluation of school accountability. Specifically, how do



score gains that may be attributed to increased test preparation on the part of most or all of the school's students (v. schools with few students able to access such preparation) differentially impact school accountability? Finally, the test performance data related to test preparation gains have been based on college-going populations, not on all students. Consequently, there is a need to determine the effects of preparation programs on all-student performance for admission tests used as statewide assessments for accountability purposes.

**Contrary Score Meanings.** A concern related to admission test use to assess statewide achievement and accountability is the issue of consistency when applied to the interpretation and meaning of results. This is a predominant concern when a test originally intended and developed for one purpose and one population now has multiple purposes for multiple populations. This issue of consistency, for example, might reflect an unsatisfactory rating for one use of the test for one purpose (as a measure of high school achievement) and a satisfactory rating for another use of the test (as a benchmark for college readiness, or as a college admission and/or placement measure), with both ratings based upon the exact same test performance.

This issue becomes clear when examining scores from the PSAE. Specifically, there appears to be a particular subset of results where disparities exist in terms of success indicators across multiple uses. For example, score distributions in 2006 (ACT & the Illinois State Board of Education, 2006) show that over 200 students whose mathematics scores were in the two lowest performing categories (Below Standards and Academic Warning) actually had ACT Mathematics scores that were at or above the ACT Mathematics College Readiness Benchmark. Similarly, nearly 1,750 students with reading scores in the two lowest performing categories actually had ACT Reading scores that were at or above the ACT Reading College Readiness Benchmark. Additionally, these same low-performing scores were at or above the suggested national ACT cutoff score guide for placement into first-year mathematics and social studies courses (ACT, 2009d).

Using the same data, a large number of Illinois students (between 15,000 and 29,000,

depending upon the institution) whose mathematics scores fell within the two lowest PSAE categories would have simultaneously fallen within the middle 50 percent of enrolled students' ACT Mathematics scores at Eastern Illinois University, Northeastern Illinois University, and/or Southern Illinois University.<sup>5</sup> Additionally, a substantial number of Illinois students (between 2,600 and 12,000, depending upon the institution) whose reading scores fell within the two lowest PSAE categories would have simultaneously been partially or fully exempt from all or part of institutional placement testing and placed into credit-bearing courses (as opposed to developmental/remedial coursework) at many of the state's largest community colleges.<sup>6</sup>

Likewise, with the MHSA, there are disparate score interpretations. A significant number of Maine students (between 2,300 and 2,600, depending upon the institution) whose reading and writing scores fell within the Partially Meets the Standards category on the 2006 MHSA administration (Measured Progress, 2007) would have also fallen within the middle 50 percent of enrolled students' SAT critical reading and writing scores at the University of Southern Maine, the University of Maine at Machias, and/or Maine Maritime Academy (see footnote 5). Additionally, about 1,350 Maine students whose reading scores fell within the Partially Meets the Standards category would have been fully exempt from institutional placement testing and placed into credit-bearing courses in at least one of the state's community colleges (e.g., Southern Maine Community College).

This issue of consistency in interpretation and meaning in terms of academic achievement and college readiness, given the multiple uses of admission tests, presents a significant and not easily reconciled chal-

---

5. Data taken from the National Center for Education Statistics College Navigator, available at <http://nces.ed.gov/collegenavigator/>

6. Illinois community colleges with very high enrollments were identified through the National Center for Education Statistics College Navigator (see footnote 5) and institutional websites from Moraine Valley Community College, College of Lake County, Triton College, Harper College, and City Colleges of Chicago - Wilbur Wright College provided the actual ACT placement score data.

lenge to state policymakers (at all levels). It is clearly difficult to explain to test takers, parents, and high schools how a test can simultaneously signify low achievement and low accountability at the secondary school level and satisfactory postsecondary preparation at the college level. Such contradictions may be more likely to occur when a test originally intended as a national college admission measure is now used to assess statewide achievement and accountability.

It is not clear how to ameliorate this contradictory situation. The cause will likely be found among an interaction of the following:

- There is a clear misalignment of standards across state P-16 educational levels.
- There is a clear misalignment between state standards and assessments at the secondary and/or postsecondary levels.
- There is lack of consistency concerning the meaning of college readiness across higher education institutions within the same state.
- The use of an admission test for purposes other than originally intended will inevitably create these kinds of disparities.

This is a major issue for state policymakers to resolve as it directly reflects upon the credibility of statewide assessment results.



## SUMMARY AND RECOMMENDATIONS

This Special Report has provided a background and set of key issues related to the use of national college admission tests to assess student achievement of state high school learning standards and high school accountability. It has discussed learning standards and assessment, provided a perspective on admission test use to assess statewide achievement and accountability, and included a number of important issues that examine the fundamentals related to this relatively new test application. This report now offers recommendations for state policymakers, practitioners, and researchers:

- **When considering the use of admission tests to evaluate achievement and accountability, state policymakers should establish both initial and long-term sets of standards and assessment goals, along with formal plans for structured outcome evaluations.**

This report has outlined a number of advantages as well as concerns associated with admission tests as measures of student achievement and school accountability. Initial and long-term standards-assessment goals, along with corresponding outcome evaluations, will help ensure that all educational stakeholders and constituencies affected by these assessments are being appropriately served.

- **When using an admission test for accountability purposes, states should complete a series of initial studies that evaluate and document the alignment of the test with the state standards. These studies should be based upon the input from a range of stakeholders.**

Federal regulations mandate that states perform alignment studies when initially implementing assessments into their NCLB accountability frameworks. States should consider expanding these requirements to include multiple judgments regarding alignment and also consider implementing more than one method when instituting alignment studies.

- **Federal mandates for an initial alignment study are important but not sufficient. State policymakers should undertake alignment studies on a scheduled basis. Whenever significant changes are made to standards and/or admission tests, alignment studies should be repeated.**

There is not only the need to initially determine the degree of standards-assessment alignment, but also to evaluate how that alignment has evolved over time. Admission tests have had infrequent, but important, changes over time. The criteria for evolving state standards and states' responsiveness to these changing needs have been far more dynamic.

- **When adopting admission tests to use as accountability assessments, states should carefully examine the potential need to augment these measures with state-designed questions, components, and/or whole assessments to ensure that all state-specific learning standards are adequately assessed.**

It is debatable that an admission test can sufficiently cover the breadth and depth of state learning standards. As state standards generally require schools to do more than prepare all students for college, states should consider augmenting college admission tests in every subject area to assure that all students are afforded the opportunity to demonstrate their full range of knowledge and skills.

- **States should undertake research to better understand the differential impacts that admission tests have on all students when these tests are used to assess achievement and accountability.**

Although considerable information exists regarding factors related to the differential performance of traditional college-bound student groups on admission tests, there has been little evaluation and research regarding reasons for the differential performance of diverse all-student populations on these same tests and the implications for these tests as measures of achievement and accountability.

- **State policymakers should evaluate the implications of two unintended outcomes when admission tests are used to assess achievement and**

**accountability: (1) the group effects of formal test preparation programs and (2) the inconsistency of score meanings in terms of high school achievement and college readiness.**

Policymakers should determine the degree of benefit that formal test preparation programs might offer various segments of the state's high school population and how differential access might relate to differential achievement and differential school accountability. Additionally, there is the need to resolve how identical results on a college admission test, used to assess statewide achievement and accountability, can signify low achievement and low accountability at the secondary school level and then commensurately indicate satisfactory postsecondary preparation at the college level (which reflects its primary purpose).

Given the growing use of admission tests to assess statewide achievement and accountability, and the need for more rigorous study of this test application, this report offers one final recommendation:

- **Independent organizations such as the Board on Testing and Assessment at the National Academies should be encouraged and supported to conduct a rigorous and thorough examination of the use of national admission tests to assess achievement and accountability at the state level.**

Despite its growth over the past decade, there has been very little independent analysis done on the use of admission tests to assess statewide achievement and accountability, in terms of appropriateness, value, validity, differential impact, ability to meet NCLB goals, and (intended and unintended) consequences. Such an examination would be a major help to policymakers, practitioners, and researchers as they consider and evaluate this test application for their own settings and constituencies.

## AUTHORS

**Dr. Richard J. Noeth** (rnoeth@mchsi.com) is an independent consultant whose clients include the College Board, Educational Testing Service (ETS), National Center for Education Statistics, Westat/National Assessment Governing Board, Pathways to College Network, and the National Association for College Admission Counseling (NACAC). He has recently completed a textbook, *Foundations of Standardized Admission Testing for NACAC*, with an online competency assessment. He is a former Vice President of Field Services and Executive Director of the SAT Program at ETS, Senior Program Officer for the Board on Testing and Assessment at the National Academies, and Director of Policy Research at ACT. His Ph.D. in Counseling Psychology is from Purdue University where he was named the tenth Distinguished Alumnus in the College of Education.

**Dr. David J. Rutkowski** (drutkows@indiana.edu) is an educational researcher who focuses on conducting policy-relevant research in the area of educational large-scale assessment and evaluation. David has consulted with international organizations, national governments, and research institutes on educational indicator development, educational assessment and the analysis thereof. Most recently, David consulted with a national Ministry of Education to develop a large-scale, national system for evaluating the quality of schools. David has published several works in peer-reviewed educational journals, encyclopedias, and edited books, and he has co-authored research projects for the United Nations Educational Scientific and Cultural Organization and the International Association for the Evaluation of Educational Achievement. David received his Master's degree in educational policy and his Ph.D. in educational policy with a research specialization in evaluation from the University of Illinois at Urbana-Champaign.

**Bridget A. Schleich** (bschleic@indiana.edu) is a graduate research assistant at the Center for Evaluation and Education Policy. She is in the School Psychology program at Indiana University working on her Education Specialist (Ed.S.) degree with a minor in counseling. Bridget received her Bachelors of Science from Northern Illinois University in Psychology.

## ACKNOWLEDGEMENTS

We would like to thank a number of people that have made this report possible. First, we would like to acknowledge the consortium of Northern Illinois High School District Superintendents, specifically, **Dr. Nancy Robb** and her predecessor **Dr. Roger Thornton** from High School District 211, for their financial support and review of this report. We would also like to thank the Director of the Center for Evaluation and Education Policy (CEEP), **Dr. Jonathan Plucker**, for his support during this project, as well as several CEEP and Indiana University faculty and staff that provided constructive feedback including: **Dr. Nathan Burroughs**, visiting research associate, **Stephen Hiller**, undergraduate research assistant, and **Laura Vosahlik**, graduate research assistant, all of CEEP; **Dr. Leslie Rutkowski**, assistant professor in Counseling and Educational Psychology; and **Dr. Ada Simmons**, Director of the Center for P-16 Research and Collaboration.

Additionally, we would like to acknowledge three external reviewers, **Meryl Bertenthal** from Indiana University, **Dr. Joan Herman** from UCLA, and **Dr. Judith Koenig** from the National Academies, all of whom provided helpful suggestions and feedback by reviewing and commenting on the initial study outline, and Meryl Bertenthal and Dr. Koenig for their review and recommendations on the preliminary draft of this report.

Finally, we would like to express our gratitude to **Terry Spradlin**, CEEP's Associate Director for Education Policy, who reviewed, coordinated, and oversaw the completion of the project.

We are very appreciative of the time, effort, and talent each of these individuals provided.

## REFERENCES

- Achieve. (2007). *Aligned expectations? A closer look at college admissions and placement tests*. Washington, DC: Author.
- Achieve & The Education Trust. (2008). *Transforming statewide high school assessment systems: A guide for state policymakers*. (Measures that Matter Series). Washington, DC: Achieve. Retrieved January 15, 2010, from <http://www.achieve.org/files/TransformingStatewideHighSchoolAssessmentSystems.pdf>
- ACT. (2003). *ACT national curriculum survey, 2002-2003*. Iowa City, IA: Author.
- ACT. (2006a). *ACT national curriculum survey, 2005-2006*. Iowa City, IA: Author.
- ACT. (2006b). *Ready for college and ready for work: Same or different?* Iowa City, IA: Author.
- ACT. (2009a). *Statewide administration of the ACT: A key component in improving student access to college and work*. Iowa City, IA: Author.
- ACT. (2009b). *ACT national curriculum survey, 2009*. Iowa City, IA: Author.
- ACT. (2009c). *ACT profile national: Graduating class 2009*. Retrieved February 10, 2010, from <http://www.act.org/news/data/09/pdf/National2009.pdf>
- ACT. (2009d). *The ACT user handbook*. Iowa City, IA: Author.
- ACT & The Illinois State Board of Education. (2006). *Prairie State Achievement Examination technical manual: 2006 testing cycle*. Springfield, IL: Illinois State Board of Education. Retrieved January 15, 2010, from [http://www.isbe.state.il.us/assessment/pdfs/2006\\_PSAE\\_tech\\_manual.pdf](http://www.isbe.state.il.us/assessment/pdfs/2006_PSAE_tech_manual.pdf)
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

- Briggs, D.C. (2009). *Preparation for college admission exams*. (A discussion paper for the National Association for College Admission Counseling). Arlington, VA: National Association for College Admission Counseling.
- Brown, R.S., & Niemi, D.N. (2007). *Investigating the alignment of high school and community college assessments in California*. (National Center Report 07-3). San Jose, CA: The National Center for Public Policy and Higher Education.
- Callan, P.M., Finney, J.E., Kirst, M.W., Usdan, M.D., & Venezia, A. (2006). *Claiming common ground: State policy-making for improving college readiness and success*. San Jose, CA: The National Center for Public Policy and Higher Education.
- College Board. (2005). *Report for the State of Maine on the alignment of the SAT and PSAT/NMSQT to the Maine Learning Results*. New York: The College Board. Retrieved January 25, 2010, from [http://maine.gov/education/mhsa/docs/sat\\_alignment\\_study.pdf](http://maine.gov/education/mhsa/docs/sat_alignment_study.pdf)
- College Board. (2009). *2009 college bound seniors: Total group profile report*. New York: The College Board. Retrieved February 9, 2010, from <http://professionals.collegeboard.com/profdownload/cbs-2009-national-TOTAL-GROUP.pdf>
- Conley, D.T. (2007). *Redefining college readiness*. (Prepared for the Bill & Melinda Gates Foundation.) Eugene, OR: University of Oregon, Educational Policy Improvement Center.
- Dossey, J., & McCrone, S.S. (2006). *External review of the Prairie State Achievement Examination Mathematics Test*. In ISBE, *Prairie state achievement examination technical manual: 2006 testing cycle*. Retrieved January 8, 2010, from [http://www.isbe.state.il.us/assessment/pdfs/2006\\_PSAE\\_tech\\_manual.pdf](http://www.isbe.state.il.us/assessment/pdfs/2006_PSAE_tech_manual.pdf)
- Gendron, S.A. (2006, July). *Letter to assistant secretary for elementary and secondary education Johnson*. Retrieved January 22, 2010, from <http://www.state.me.us/education/nclb/reconsideration/reconsiderationstatus073106.pdf>
- Hamilton, S.L., Stecher, B.M., Marsh, J.A., McCombs, J.S., Robyn, A., Russell, J., et al. (2007). *Standards-based accountability under No Child Left Behind: Experiences of teachers and administrators in three states*. Santa Monica, CA: RAND Corporation.
- Illinois State Board of Education. (2008). *PSAE modification*. Retrieved February 10, 2010, from <http://www.isbe.net/assessment/listserv/2008/mar14.htm>
- Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education*. Washington, DC: Science Directorate, American Psychological Association.
- Lawrence, I., Rigol, G.W., Van Essen, T., & Jackson, C.A. (2002). *A historical perspective on the SAT: 1926-2001*. (College Board Research Report No. 2002-7). New York: College Entrance Examination Board.
- Linn, R.L. (2005). *Issues in the design of accountability systems*. *CSE Technical Report 650*. Los Angeles: UCLA, Center for Research on Evaluation, Standards, and Student Testing.
- Linn, R.L., & Miller, M.D. (2005). *Measurement and assessment in teaching* (9<sup>th</sup> ed.). New York: Pearson.
- Mattern, K.D., Patterson, B.F., Shaw, E.J., Kobrin, J.L., & Barbuti, S.M. (2008). *Differential validity and prediction of the SAT*. (College Board Research Report No. 2008-4). New York: The College Board.
- Measured Progress. (2007). *Maine Comprehensive Assessment System: 2006-2007 technical manual part two - Maine High School Assessment*. Dover, NH: Author. Retrieved January 15, 2010, from <http://www.maine.gov/education/mhsa/docs/07techmanual.pdf>
- Mitchell, R. (2006). *Standardized and high-stakes tests*. Alexandria, VA: The Center for Public Education.
- National Association for College Admission Counseling. (2008). *Report of the commission of the use of standardized tests in undergraduate admissions*. Arlington, VA: Author.
- Noeth, R.J. (2009). *Foundations of standardized admission testing*. Arlington, VA: National Association for College Admission Counseling.
- Ogle, D., & Hunter, K. (2006). *External review of the Prairie State Achievement Examination reading and writing tests*. In ISBE, *Prairie state achievement examination technical manual: 2006 testing cycle*. Retrieved January 8, 2010, from [http://www.isbe.state.il.us/assessment/pdfs/2006\\_PSAE\\_tech\\_manual.pdf](http://www.isbe.state.il.us/assessment/pdfs/2006_PSAE_tech_manual.pdf)
- Payne, D.A. (1997). *Applied educational assessment*. Boston: Wadsworth.
- Schafer, W.D. (2003). A state perspective on multiple measures in school accountability. *Educational Measurement: Issues & Practice*, 22(2), 27-31.
- Shepard, L., Hannaway, J., & Baker, E. (2009). *Standards assessment and accountability*. Washington DC: National Academy of Education.
- Southern Regional Education Board. (2007). *High school to college and careers: Aligning state policies*. Atlanta, GA: Author.
- U.S. Department of Education. (2003). *Standards and assessment: Non-regulatory guidance*. Retrieved February 10, 2010, from <http://www2.ed.gov/policy/elsec/guid/saaguidance03.doc>
- Webb, N. L. (2006). *Alignment analysis of secondary mathematics standards and the SAT reasoning test*. Augusta, ME: Maine Department of Education.
- Zwick, R. (2007). *College admission testing*. Arlington, VA: National Association for College Admission Counseling.

---

## WEB RESOURCES

### **Achieve**

<http://www.achieve.org>

### **ACT**

<http://www.act.org>

### **Center for Evaluation & Education Policy, Education Policy Briefs**

<http://www.ceep.indiana.edu/pub.shtml#ed>

### **College Board**

<http://www.collegeboard.com>

### **Illinois Prairie State Achievement Examination (PSAE)**

<http://www.isbe.state.il.us/assessment/psae.htm>

### **Maine High School Assessment (MHSA)**

<http://www.state.me.us/education/mhsa/index.htm>

### **U.S. Department of Education, No Child Left Behind**

<http://www2.ed.gov/nclb/landing.jhtml>

---

CEEP Special Reports are executive edited by Jonathan A. Plucker, Ph.D. and published by the

## **Center for Evaluation & Education Policy**

Indiana University  
1900 East Tenth Street  
Bloomington, IN 47406-7512  
812-855-4438

More about the Center for Evaluation & Education Policy  
and our publications can be found at our website:  
<http://ceep.indiana.edu>