

Race to the Top Public and Expert Input Meeting
Comments on the Design and Development of a Common Core High
School Assessment

November 13, 2009

Boston, MA

Wayne J. Camara
Vice President, Research and Development
The College Board
45 Columbus Ave.
New York, NY 10023 (212) 713-8069
wcamara@collegeboard.org



I am pleased to have this opportunity to offer some brief comments on the design and development of a high quality summative assessment system that can be used by consortia of states to measure college and career readiness. The College Board will submit more detailed written comments by the December 2nd deadline.

Today I will restrict my comments to a few essential concepts concerning the high school assessment.

First, I recommend that the Department of Education formally recognize the **Standards for Educational and Psychological Testing** (AERA, APA, NCME, 1999) as providing definitive professional guidance on the development and use of any assessments related to this initiative. In the request for input, the Department has called for “high quality summative assessments” that are based on “best practices in assessment.” In addition, the request appropriately requires such assessments provide evidence relating to their validity, reliability, and fairness. The **Standards for Educational and Psychological Testing** have served as the definitive source for assessment professionals across a variety of applications (e.g., education, employment, licensure, psychological) and delineate the appropriate types of evidence that are required to support statements by test publishers and users concerning these and other claims (e.g., comparability, use of cut scores). The Department of Education should ensure that any proposed summative assessment appropriate addresses these standards and that a technical oversight group is in place to review the proposed use(s) and evidence.

The Department encourages assessments that are innovative. Some parties may claim that adhering to established standards for psychometric quality, validity and fairness will prevent innovation. This argument is specious and suggests that innovation is inconsistent with quality and best practice. The Standards address this issue and note that “the applicability of the Standards to an evaluation device or method is not altered by the label applied to it...the degree to which stimulus materials are standardized...or the type of response format (p. 3).” It is essential that the Department provide support and resources to ensure high quality assessments are in place¹.

The Standards recognize that new assessments may not initially have all the documentation and evidence required to support inferences about validity. However, such evidence can be gathered overtime and should be equally required of any assessment or accountability system.

Second, the goals and intended purposes of this new assessment will be best served through an integrated assessment system that includes summative, interim and formative elements. Summative assessments can best provide useful information to students, parents, and schools on college and career readiness. Valid and reliable inferences can be produced for student and school level decisions. This information may also inform other decisions such as course placement, teaching and learning, and student growth or changes in achievement if additional information is incorporated into the system beyond that collected during a single summative assessment. For example, a math test administered in 11th grade may not be the more precise way to predict how well a student will perform in a college math class some 18 months into the future. This is especially true when students score close to the cut point or when they fail to

continue to take a math course in their senior year. Interim assessments can provide snapshots of how students are doing in mastering skills or providing more in-depth analysis of student weaknesses at a point in time. The formative components of such an integrated system can complement the summative and interim assessments and provide instructionally actionable information to schools and districts. A carefully designed integrated system is needed to ensure all components are complementary and consistent. Formative and interim assessments could utilize a common bank of assessment tasks and scoring rubrics available for teacher use.

Third, specifying the intended purposes of the summative assessment is the first step in designing a quality assessment. At least nine purposes have initially been mentioned in the Department's call for inputs for the summative assessment:

1. To inform teaching and learning
2. To determine school effectiveness
3. To determine teacher and principle effectiveness
4. To determine student readiness for college and careers
5. To determine if a student is on track for college and career readiness
6. To measure student growth or change in achievement
7. To determine high school graduation
8. To determine college course placements
9. To inform college admissions

A single summative assessment or assessment system cannot possibly serve all of these uses equally well. There are tensions between many of these uses and constraints that impose significant operational requirements for other uses. For example, summative assessments are not designed to provide instructionally rich and actionable information. Typically results are not available till the end of a school year while diagnostic information is needed throughout the year. Another constraint and conflict exists between the desire for innovative assessments that take advantage of technology and the use of the same assessments for very high stakes individual decisions. Many state assessments are delivered by computer, but they have done so by permitting schools to administer the same form (and or items) over an extended testing window. There are simply not enough computers in schools to administer the same test to all 8th graders in a state on a single date (or 3-4 different dates). School calendars also vary greatly within a state and flexibility in administration is required to accommodate local demands. Contrast this requirement with the security demands placed on tests used for college admissions, college credit and college placement. National testing programs have extensive procedures to ensure the security of test content and results for such high stakes programs. The same items and forms cannot be administered over an extended window without greatly compromising security. In addition, the number of item pools and items required to maintain security of adaptive programs that offer the same level of flexibility for administrative dates would be cost prohibitive. These and other trade-offs need to be considered in determining the final requirements and purposes for an assessment system. The Department of Education should identify a limited number of desired uses for a summative assessment system. In each instance, the consortium of states should then describe the types of evidence that will be used to support the validity of inferences that will be made for each purpose.

Fourth, the system is designed to provide students and schools with valid and reliable indicators on college and career readiness. These as other educational outcomes, are best measured with multiple measures. In a forthcoming research report, the College Board will propose a method to estimate college readiness of students in a school or state based on three academic metrics: (1) SAT scores; (2) high school grades; and (3) a new measure of high school academic intensity that considers the number of courses completed, the highest level of courses completed and the rigor of course taken (e.g., honors, dual enrollment, AP). David Conley and others have noted that there are other relevant factors in determining college readiness, yet any accountability system that is based solely on test scores would ignore consistent findings from research that demonstrate the importance of examining the rigor of courses taken and student achievement in those courses. Ultimately, the validity evidence to support college readiness assessments must include predictive evidence of the relationship between the assessment results and essential outcomes of college success (e.g., college going, placement into credit bearing courses without remediation, academic grades, time to degree, persistence). Current state assessment systems primarily rely on content validity evidence with little focus on their relationship to future performance. Expert judgment of content is an important form of evidence, yet predictive and statistical evidence is essential in evaluating the efficacy of assessments used to determine college and career readiness.

The final topic that will be discussed is innovation in assessment and the College Board's written testimony will discuss this issue and propose more specific models for a high quality assessment system. Innovation can be realized most efficiently in a large scale testing program if it is delivered exclusively on computer. Innovative item types, extended performances and different response formats can be more efficiently captured and scored with the use of technology. Innovation in large scale assessment has been hampered by the requirement to produce comparable forms on paper. If the assessment is administered solely on computer (with the exception of paper administration as a special accommodation) it will be easier to introduce new item types such as simulations, scenario-based tasks, or performance tasks. Ideally such tasks in the summative assessment can largely be scored by computer to increase efficiency and reduce turnaround time. Teacher scoring of formative and/or interim assessments can be best utilized in a distributed scoring network or through an audit function.

Many of the emerging skills contained in the draft common core college and career readiness standards can likely not be measured with paper based assessments alone. Maintaining parallel paper and computer systems would likely limit innovation and the range of emerging skills that could be measured. This is another example of the trade-offs that must be considered in the final design of assessment systems that will be proposed by state consortiums.

Another option is to incorporate results from interim assessments or actual student performances that occur throughout the year into the summative assessment score. Currently, summative assessments are based on what a student does at the end of the year on a single test date, while some high performing nations have incorporated student performance at several different points in time into their summative assessment. Results from interim assessments or tasks completed during the year or student performance on a highly structured in-class or out-of-class assignment (e.g., research paper, literary report, laboratory report, presentation) that is scored by teachers using a detailed scoring rubric could be incorporated into the results of summative assessments.

Clearly such models present operational challenges in terms of security and when students transfer into a school midway through the year, yet such models could increase the instructional relevance of assessments and work for the vast majority of students. These are other elements will be discussed more fully in the final written comments due next month.

In closing, having a common core assessment component that is employed across states is essential to having comparable results and indicators of students college and career readiness. States may be expected to augment the common core summative assessment with additional items and tasks that supplement the core assessment. In addition, separate proficiency levels may be developed at the state and national level, but maintaining common components of an assessment is essential if we desire to increase the level of objective and comparable data on student and school performance.