## Addressing the Variability of Learners in Common Core-Aligned Assessments: *Policies, Practices, and Universal Design for Learning*

## Introduction

Monitoring and assessing the achievement of students is a key component of the curriculum at any instructional level. Because these measures have implications for important educational decision-making, it is essential that the measures are (1) accurate, (2) useful for subsequent educational planning and (3) sufficiently timely to benefit each student. The creation of the Common Core aligned PARRC and SBAC assessments has captured national attention in two primary areas. First, even though formative assessment procedures are stated components of both of these assessment design and development consortia, their primary emphasis has been on the establishment of large-scale summative instruments, which are likely to be implemented in many states as high stakes tests, used as a determinant for grade-level promotion and high school graduation. Second, these summative measures are designed, from the outset, to be delivered in digital formats, a different medium from the paper and pencil versions with which students and educators have more familiarity.

Viewing these challenges within the framework of Universal Design for Learning (UDL) and its emphasis on providing students with multiple means of engagement, representation, and action and expression, some key issues emerge: First, instruction and assessment share dependencies within the curricular cycle, and expansions or constrictions in one area affects the other. Second, the real-time, more formative, achievement monitoring that is increasingly a component of digital curriculum resources and the systems that deliver them, when combined with learning analytics and large data set trend analysis, provides previously unavailable opportunities for applying pedagogical interventions at the point of instruction. These data-driven monitoring capabilities, the subject of considerable attention and investment by the United States Department of Education's Office of Educational Technology and highlighted in the 2010 *National Education Technology Plan* (2010), promote the benefits of embedded, real-time, versus extrinsic approaches to assessment.

With the above issues and the framework of Universal Design for Learning in mind, CAST has identified five critical factors that should be addressed from the outset by PARCC and SBAC when creating assessments, both formative and summative:

1. Move away from apparent exclusive focus on summative measures and prioritize formative assessments as part of the assessment instruction cycle.

2. Capitalize on the use of technology-based assessments to ensure that the benefits—flexibility, real-time monitoring of student progress and the promotion of access for all students—are realized.

3. Consider the impact of assessment on classroom instruction in order to facilitate rather than constrain the modification of instruction based on student performance.

4. Be mindful of the potential negative effects of *computer-adaptive testing* (CAT) on all subgroups, including students with disabilities and English Language Learners.

5. Ensure—for all students—accuracy, reliability, and precision with respect to intended constructs.

## Critical Factors

**(1) Move away from apparent exclusive focus on summative measures and prioritize formative assessments as part of the assessment instruction cycle.**

- In contrast to summative assessment procedures, which provide a single snapshot of student performance at the end of an instructional episode, formative assessment allows educators to evaluate student understanding of knowledge and skills in an ongoing and embedded manner in preparation for college and career-ready standards.

- It is important for the common core assessment consortia to prioritize and address formative assessment in a meaningful way rather than repurposing large-scale assessment items for occasional interim summative measurement. Referring to mastery measurements as formative is misleading.

- Teachers, all students (including students in the 1%, students previously in the 2%, and those considered gifted and talented), administrators, and parents benefit from the data collected in well-designed formative assessment.

- The formative assessment process provides information about performance during the instructional episode so that modifications, changes, and alterations in instruction can be made to support achievement toward the instructional goals.

- All assessment data collected by states should be used to inform and improve learning and instructional practices for all students.

- Well-developed and implemented formative assessments can lead to improvements in each learner's attention to and analysis of his/her own learning process and products.

- Without established and well implemented formative assessment procedures, educators, students, and parents may not be adequately informed about progress toward a goal—in other words, they may not be informed until after it is too late to support or change instruction. For this reason, CAST supports formative assessment, specifically that of progress monitoring.

**(2) Capitalize on the use of technology-based assessments to ensure that the benefits— flexibility, real-time monitoring of student progress and the promotion of access for all students—are realized.**

- Digitally based assessments have the capability and flexibility to facilitate access to the assessment and to the general education curriculum for students with sensory, physical, and learning disabilities. The development and administration of such assessments should be designed and implemented to ensure that they are effectively facilitating access.

- Policies and procedures with respect to the use of assistive technology (AT) during assessment should ensure that they do not impede the availability of these

supports for students who need them. Technical standards for interoperability between Common Core assessments and AT devices need to be established. These standards could be used as criteria against which a wide range of assistive technologies could be evaluated.

- In cases where construct validity is not likely to be violated, the assessment consortia and state implementers should not limit the use of AT by authorizing only specific assistive technologies in the assessment process. Denying students the opportunity to use technology and AT will not help in the preparation for college and careers.

- Any process used for authorizing appropriate access and accommodations for individual students, including the development of student profiles, should include clearly established protocols, procedures, and training. In addition, those making accommodation decisions for students with disabilities, including those previously designated to participate in the modified assessment based on modified achievement standards (2%), must be adequately trained.

- It is important for states to consider the participation needs of students from all disability categories in the assessment design to help ensure that appropriate navigation and access is available throughout the entire assessment (e.g. single switch technology for students with physical disabilities). Since the major goal of the Common Core aligned assessments is to provide stakeholders with reliable and valid information that accurately describes and predicts students' readiness for career and college, the decision to exclude specific disability categories such as blind and visually impaired students is disturbing. For this reason, CAST believes that students from *all* disability categories should be included in field tests and validation studies of assessments.

**(3) Consider the impact of assessment on classroom instruction in order to facilitate rather than constrain the modification of instruction based on student performance.**

- Best practice suggests that assessment accommodations align with those accommodations that the student receives during classroom instruction.

- There is a danger of overly restrictive assessment policies and procedures driving instructional practices, including materials and tools (e.g., accessible instructional materials), used by students in the classroom. In particular, limited assessment practices could adversely impact the instructional decision-making process of the IEP team.

- Schools and/or teachers may not allow certain accommodations for instruction because these accommodations are not allowed on the assessment—for example, a state was not able to provide computer based writing tests and therefore determined that all writing instruction in classrooms should be using paper and pencil in order to parallel the annual high stakes assessment.

- Limited assessment practices can potentially create a conflict between the use of technology in instruction and availability on assessments, ultimately preventing schools from becoming more innovative in their use of multimedia tools to support *all* learners.

- All accommodations and supports provided during assessment need to be taught and practiced prior to use. Additionally, use of accommodations and supports should be made an essential component of training for teachers/administrators prior to assessment administration.

- If accommodations on assessments are inappropriately limited, these policies may inadvertently restrict the number of students who are found eligible to receive and benefit from the same accommodations during classroom instruction, in violation of their rights under IDEA and Section 504. Because of the potential impact of assessment on instruction, CAST firmly believes that the determination of assessment accommodations should be carefully addressed.

**(4) Be mindful of the potential negative effects of *computer-adaptive testing* (CAT) on all subgroups, including students with disabilities and English Language Learners.**

- There is a lack of research on the accuracy and viability of CAT on the various categories of students with disabilities (Laitusis et al., 2011); the majority of benefits for students with disabilities ascribed to CAT appear to be based on assumptions unsupported by existing research data.

- A down-leveling of test items following an item failure could result in the presentation of out-of-level items based on standards from a lower grade. This could render the assessment out of compliance with the ESEA requirement to measure student performance against the expectations for a student's grade level (Way, 2006; U.S. Department of Education, 2007). Such a result could also have the effect of violating the student's rights under IDEA and Section 504. Research suggests that maintaining alignment with content standards may be more successful if the adaptation occurs at the testlet/subtest level, rather than at the item level (Folk & Smith, 2002).

- Students with uneven skill sets may fail basic items and never have the opportunity to exhibit skills on higher-level tasks; this is particularly relevant to various students with disabilities who may exhibit idiosyncratic and uneven academic skills (Thurlow, et al., 2010; Almond, et al., 2010; Kingsbury & Houser, 2007).

- CAT approaches are reported to be efficient and accurate when item responses are limited to multiple choice and short answers (Way, 2006), while the accuracy and efficiency of more varied response types may pose significant challenges to adaptive algorithms, and hence to validity.

- The majority of CAT systems deployed to date may not allow or may significantly restrict a student's ability to return to a previous item to review or change a response (Way, 2006), further narrowing the range of test-taking strategies a student may employ. Some solutions to the application of a review and change strategy for CAT have emerged (Yen, 2012; Papanastasiou & Reckase, 2007).

- For all these reasons, CAST believes that computer adaptive testing for all students must be carefully constructed and monitored.

**(5) Ensure—for all students—accuracy, reliability, and precision with respect to intended constructs.**

- The item and task development process of the Common Core aligned assessments should ensure precision with respect to the identification of intended constructs associated with individual assessment items. CAST believes that without this precision, items or tasks will measure construct irrelevant information for certain students and, as a result, the inferences that are drawn from the assessment scores for these students will be invalid, in violation of students' rights under IDEA and Section 504.

- With respect to reading, precision of item constructs will allow a subskill such as decoding to be measured separately from higher level reading comprehension. With today's widely available technologies, students can independently demonstrate achievement of high levels of reading comprehension without having to decode specific elements of text. This logic applies to other content areas as well, for example basic calculation in contrast to higher level mathematical reasoning skills.

Sincerely,

Peggy Coyne, EdD, Research Scientist
Tracey E. Hall, PhD, Senior Research Scientist
Chuck Hitchcock, MEd, Chief of Policy and Technology
Richard Jackson, EdD, Senior Research Scientist/Associate Professor, Boston College
Joanne Karger, JD, EdD, Research Scientist/Policy Analyst
Elizabeth Murray, ScD, Senior Research Scientist/Instructional Designer
Kristin Robinson, M. Phil, MA, Instructional Designer and Research Associate
David H. Rose, EdD, Chief Education Officer and Founder
Skip Stahl, MS, Senior Policy Analyst
Sherri Wilcauskas, MA, Senior Development Officer
Joy Zabala, EdD, Director of Technical Assistance

# References

Almond, P., Winter, P., Cameto, R., Russell, M., Sato, E., Clarke-Midura, J., ... Lazarus, S. (2010). Technology-enabled and universally designed assessment: Considering access in measuring the achievement of students with disabilities—A foundation for research. *Journal of Technology, Learning, and Assessment,10(5).* Retrieved from http://ejournals.bc.edu/ojs/index.php/jtla/article/view/1605

Folk, V. G. & Smith, R. L. (2002). Models for delivery of CBTs. In C. Mills, M. Potenza, J. Fremer, & W. Ward (Eds.), *Computer-based testing: Building the foundation for future assessments* (pp. 41−66). Mahwah, NJ: Erlbaum.

Kingsbury, G. G. & Houser, R. L. (2007). ICAT: An adaptive testing procedure to allow the identification of idiosyncratic knowledge patterns. In D. J. Weiss (Ed.). *Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing.*

Laitusis, C. C., Buzick, H. M., Cook, L., & Stone, E. (2011). Adaptive Testing Options for Accountability Assessments. In M. Russell & M. Kavanaugh (Eds.), *Assessing Students in the Margins: Challenges, Strategies, and Techniques*. Charlotte, NC: Information Age Publishing.

Papanastasiou, E. C., & Reckase, M. D. (2007). A "rearrangement procedure" for scoring adaptive tests with review options. *International Journal of Testing, 7*(4), 387-407.

Thurlow, M., Lazarus, S. S., Albus, D., & Hodgson, J. (2010). *Computer-based testing: Practices and considerations* (Synthesis Report 78). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

U.S. Department of Education. (2007). *Standards and assessments peer review guidance: Information and examples for meeting requirements of the No Child Left Behind Act of 2001*. Retrieved from www.ed.gov/policy/elsec/guid/saaprguidance.doc

U.S. Department of Education, Office of Educational Technology. (2010). *Transforming American education: Learning powered by technology*. National Education Technology Plan 2010. Washington, DC: Author. Retrieved from http://www.ed.gov/sites/default/files/netp2010.pdf.

Way, W. D. (2006). *Practical questions in introducing computerized adaptive testing for K–12 assessment. PEM Research Reports*. Iowa City, IA: Pearson Educational Measurement.

Yen, Y. C., Ho, R. G., Liao, W. W., & Chen, L. J. (2012). Reducing the Impact of Inappropriate Items on Reviewable Computerized Adaptive Testing. *Educational Technology & Society 15*(2), 231-243.