Meeting the Needs of Special Education Students: Recommendations for the Race to the Top Consortia and States

MARTHA L. THURLOW, RACHEL F. QUENEMOEN, & SHERYL S. LAZARUS
Executive Summary

Special education students in the United States make up 13 percent of public school enrollment. The majority of these students can meet the same achievement standards as other students if they are given access to the same content as their typical peers and are provided specially designed instruction, supports, and accommodations when needed.

Too often, states have had to retrofit their assessments for special education students. Participation policies have varied across states, and special education students’ needs for assessment accommodations have challenged states. Now, with agreement on the Common Core State Standards, variations across states can be eliminated and common accommodations policies can be developed. Assessments can be designed from the beginning with consideration of all students, including special education students.

The Race to the Top assessment consortia also will want to consider the implications for special education students of computer-based testing and formative and interim assessment models. The consortia are focused on the regular assessment, but to ensure a coherent assessment system they also will want to regularly communicate with the National Center and State Collaborative and the Dynamic Learning Maps Alternate Assessment Consortium, which are developing alternate assessments based on alternate achievement standards for students with significant cognitive disabilities. And, although the goal is to ensure inclusive, innovative assessments that produce valid results for all students, the consortia will want to prompt districts and schools to address...
widespread instructional issues for all students, including those with disabilities.

This paper identifies several actions for the Race to the Top assessment consortia to take to meet the needs of special education students. They are consistent with standards and principles for assessments, and they reflect evolving research and development activities directed toward supporting better assessments for every student:

1. Develop a set of common accommodation policies for the Race to the Top assessments
2. Follow accessibility principles in development, field testing, and implementation
3. Ensure that the design of computer-based tests is appropriate for special education students as well as other students
4. Develop formative and interim assessments to ensure inclusion of special education students in grade-level curricula focused on accelerated learning
5. Communicate and coordinate with the alternate assessment consortia

These five recommendations require communication with expert stakeholders and a commitment to grade-level content to ensure that special education students, like other students, complete their school careers ready for college or a career. Careful planning will ensure that the Race to the Top assessment consortia build on what is already known rather than moving forward in a way that may make tests less accessible and less valid for some special education students. The time is right to build innovative assessments that are best for all U.S. students rather than for a subset of students.
Meeting the Needs of Special Education Students: Recommendations for the Race-to-the-Top Consortia and States

U.S. states have been working for more than three decades toward including all students in their education systems. This commitment has permeated both educational service provision and the approaches used to evaluate educational systems’ success in meeting the needs of all students. We have learned that students with disabilities should not be pitied or protected from the same high expectations we have for other students. Nor should they be excluded from the assessments that tell us how we are doing in making sure that they meet those expectations.

A commitment to the inclusion of students who receive special education services accompanied the standards-based education movement that started in the early 1990s. That commitment continued, reinforced by the requirements of federal laws (first the Elementary and Secondary Education Act and then the Individuals with Disabilities Education Act), as states adopted their own standards and then the new Common Core State Standards in English language arts and mathematics. This commitment is now challenged by questions about how best to include special education students as states move toward innovative approaches to assessment through Race to the Top (RTTT) funding.

As the Race to the Top assessment consortia and states explore challenges in meeting the needs of special education students and work to develop shared solutions, it is important that they start with common understandings of who these students are. They must commit to successfully including these students in common assessments based on the Common Core Standards, no matter what the assessment or how innovative it may be. This paper describes what has been gleaned over the past 20 years about who these students are and what they need to learn successfully and to demonstrate their learning. The paper identifies challenges that states must address in developing assessment systems that include these students. In it are suggestions for specific ways to address these challenges, toward the goal of developing RTTT assessments for all students, including all students with disabilities who receive special education services or who have

The vast majority of special education students (80–85%) can meet the same achievement standards as other students if they are given specially designed instruction, appropriate access, supports, and accommodations, as required by IDEA.
Section 504 accommodation plans (developed for students with disabilities who need accommodations but do not necessarily need special education services).

Who Special Education Students Are—Implications for Instruction and Assessment

Students with disabilities who receive special education services as required by the Individuals with Disabilities Education Act (IDEA) currently make up 13 percent of public school enrollment, with percentages in states varying from 10 percent to 19 percent. Special education students are disproportionately poor, minority, and English language learners.

The vast majority of special education students (80-85 percent) can meet the same achievement standards as other students if they are given specially designed instruction, appropriate access, supports, and accommodations, as required by IDEA. Figure 1 displays the categorical distribution of special education students. Although disability category is not the best indicator of students’ strengths and needs, it does serve as a proxy for understanding that only a small percentage of special education students have a disability that may require different achievement standards. This small group would include some, but not all, students with intellectual impairments, autism, or multiple disabilities, plus potentially a few students with other disability labels. In addition, researchers have learned that even special education students who require different achievement standards can—when given high-quality instruction in the grade-level curriculum—do far more than has been seen in the past.

FIGURE 1: DISTRIBUTION OF DISABILITY CATEGORIES IN 2008–09

Reprinted with permission from the National Center for Learning Disabilities.2
Simply put, it is irresponsible to assume that because a student receives special education services, that student is a low performer who cannot learn. Rather, the goal should be to ensure that special education students progress through school successfully, with access to the same content as their typical peers to be ready for college or a career. Also, assessments must be designed from the beginning to ensure that the ways in which students are asked to demonstrate their knowledge and skills do not create barriers related to disabilities that have nothing to do with what is being measured.

**Participation and Performance of Special Education Students in Pre-RTTT Assessments**

Students with disabilities have benefited in many ways from the U.S. focus on standards and assessments. After decades of being excluded from state and district assessment systems, their participation in state assessments has increased from no more than 10 percent of students with disabilities participating in the early 1990s to an average of 99 percent at the elementary level, 98 percent at the middle school level, and 95 percent at the high school level in 2007-08. These increases are due in large part to participation requirements in the Elementary and Secondary Education Act (ESEA) and IDEA.

We also are seeing evidence of improvements in the academic performance of students with disabilities. Some of this evidence comes from trends in their performance on the National Assessment of Educational Progress, in which the grade 4 performance of students with disabilities in 2005 was significantly higher than it was in 1998, 2000, 2002, and 2003 (figure 2).

**FIGURE 2: GRADE 4 PERFORMANCE OF STUDENTS WITH AND WITHOUT DISABILITIES ON NAEP READING ASSESSMENT**

![Figure 2: Grade 4 Performance of Students with and without Disabilities on NAEP Reading Assessment](source: National Center for Learning Disabilities, 2007. Reprinted with permission of the National Center for Learning Disabilities.)

*Significantly different from 2005
Although large gaps still exist in performance between special education students and their peers without disabilities, there is now a better understanding about these students, their opportunities to learn, and what can be expected of them. Much has also been learned about what needs to change in their instruction, access to the curriculum, and in assessments in order to first see their achievement increase dramatically, and then to capture that achievement on inclusive assessments.

These findings are all important for the RTTT assessment consortia and states to consider when developing assessments that best meet the needs of all students, including special education students. The RTTT assessment consortia and states can ensure that new assessments build on, rather than step back from, what has already been learned. Too often, assessments have been retrofitted for special education students. The RTTT assessment consortia and states now have an opportunity to consider such students’ needs from the start.

FIGURE 3: EIGHT-YEAR TRENDS OF THE PERCENTAGE OF ELEMENTARY STUDENTS WITH DISABILITIES WHO ACHIEVED PROFICIENCY ON STATEWIDE READING EXAMS

Source: Figure 23 in Good News and Bad News in Disaggregated Subgroup Reporting to the Public on 2005-2006 Assessment Results. Reprinted with permission from the National Center on Educational Outcomes.
Common Consortia Accommodations Policies

State participation policies have varied dramatically over the years, making it difficult to compare results across states and across time. Participation policies guide decisions about how students with disabilities participate in assessment systems. Accommodations are changes in the materials or procedures used for instruction and assessment. States have been challenged by many special education students’ need for assessment accommodations. This is evident in the varying policies on accommodations across states (see figure 4 for states’ read aloud questions accommodation policy on tests of reading) and the varying percentages of special education students receiving accommodations (figure 5). With agreement on the Common Core Standards, it is time to reach consensus on accommodation policies.

FIGURE 4: 2008–2009 POLICIES ON USE OF THE READ ALOUD QUESTIONS ACCOMMODATION FOR TESTS OF READING
Identifying common accommodation policies on which all states in the RTTT assessment consortia agree will require careful explication of the content standards, to define precisely what content is to be assessed and what content is not the focus of assessment. To the extent that content targets are intertwined, such as decoding and understanding text, approaches should be developed that separate skills that serve as gatekeepers (such as decoding) to understanding a student’s other well-developed skills and knowledge.
Too often, students with disabilities are penalized because their disabilities create barriers to proper assessment. For example, phonemic awareness and decoding are separate skills from understanding and drawing inferences from text. A student may be a poor decoder but may be able to comprehend text well when using text-to-speech technology or read-aloud accommodations. Such accommodations are commonly used by adults with reading disabilities and enable them to succeed in every field of endeavor. Students’ knowledge cannot be measured if it is not recognized when artificial barriers prevent them from demonstrating their knowledge and skills.

Research on the effects of accommodations is increasing. One of the critical findings of that research is that careful definition of the content to be taught and assessed—what assessments are really trying to measure—is of critical importance. With explicit definitions of content targets, assessments that measure what is intended can more easily be developed. Further, we can figure out how to increase the accessibility of the assessments and the accommodations that some students may still need. Clear, grade-level content definitions are critical to the development of assessments appropriate for the greatest number of students, including special education students. Ensuring that common standards have addressed accessibility concerns does not mean lowering the standards. It does mean, for example, providing a way for students who cannot hear to demonstrate their “listening” skills; for students who cannot see to demonstrate their “viewing” skills; and for students who cannot decode to demonstrate their comprehension skills in reading.

**Universal Design for Innovative Assessments**

Universal design principles address policies and practices that are intended to improve access to learning and assessments for all students. Universal design applied from the beginning of test development through to the point when students engage in assessments is important because some assessment designs pose barriers that bar students with disabilities from showing what they know. Universal design techniques may reduce the need for accommodations and alternate assessments and may provide states with more cost-effective assessments. Further, tests designed with universal design principles can provide educators with more valid inferences about the achievement levels of students with disabilities, as well as the achievement of their peers without disabilities.

Universal design approaches recognize that students with disabilities are general education students first. The major implication of this approach is that assessments should be better designed from the beginning with all students in mind but that they should not be different for most students with disabilities. Whether these are called universally designed assessments or accessible assessments, it is critical that all students be considered during the development process and that what each assessment is intended to measure be clearly defined, as well as how that content can be measured for all students, including those whose disabilities create challenges for showing what they know and can do. Tests administered on computers are not universally designed assessments. There is the potential that computer-based tests may be less accessible than well-designed paper-pencil tests if care is not taken to consider the needs of all students.

**Computer-based Testing**

Computer-based testing shows promise for increasing the accessibility of assessments for all students, including those receiving special education services. Computer-based tests can be developed in a way that embeds what would be “accommodations” when the test is paper-based so that they are simply features of the test design.
rather than accommodations. These might include, for example, navigation via touch screens, sip-and-puff devices, or switches as well as the standard computer mouse. Text can be read aloud or signed on the computer; color contrasts and masking features can be embedded; and a host of other features can be used to ensure that all students are assessed in a way that produces valid results. In addition, changes that continue to be considered as accommodations can be tracked.8

Though the possibilities for dramatically increasing the accessibility of assessments in a computer-based assessment system are significant, accessibility issues specific to special education students continue to present challenges in computer-based testing.9 Challenges include less access to computers and less experience with keyboarding for some students with disabilities compared to other students. Other difficulties may be related to certain disability characteristics. For example, some computer-based tests may require more working memory than is needed for a similar paper and pencil test. Computer-based tests can be difficult for students whose fine motor skills are poor or who have severe visual impairments. English language learners with disabilities may face challenges if a digitalized or human text-to-speech voice uses an unfamiliar dialect because dialects may not be easily perceived and understood by these students.

Computer-based systems should be developed to be as transparent as possible about the content on which students are assessed and the ways in which the content is assessed. When assessment results are to be used for system accountability, they must be standards-based, even in the name of measuring growth, or they will risk sacrificing the goal of closing well-documented achievement gaps for statistical convenience. Computer-based systems should not revert to an out-of-level testing approach in which students are tested on lower-grade-level content. Any adaptive computer-based assessments must be restricted to grade-level content and must be transparent enough to know when a student is inaccurately measured because of poor basic skills but good higher level skills. Without these constraints, the “any progress will do” approach that characterized Title I and IDEA prior to 1994 will recur, with little to no reduction in the achievement gap.

**Formative and Interim Assessments**

Assessments used to measure students more frequently during the course of the year are an important part of a comprehensive assessment system. These assessments must be designed with the same attention to access as have been required in state summative assessments.

Just as the development of new innovative summative assessments through the RTTT assessment consortia provides the opportunity to ensure that all students, including special education students, are considered in the initial design and development stages, so, too, must that opportunity be taken for formative and interim assessments. In addition to considering the design of these assessments and the use of accommodations, developers must address content issues. For example, the argument that formative assessments must be sensitive to a student’s instructional level must be weighed in light of the need to focus the student’s instruction on grade-level content.

In the past, there was a tendency for teachers to identify missing skills or knowledge, often those typically taught in earlier grades, and then institute remediation until those missing skills were mastered. Research on learning progressions in recent years has shown that many teachers did not know which missing skills were essential to ensure that a student remained on target to “catch up” in the essential skills and knowledge, and thus endless remediation often resulted in a situation where the student could never catch up. Many states and districts have realized that teachers need support in un-
derstanding the different ways that students can “make up for lost time”—the varied content pathways teachers can use to scaffold students’ learning to lift them around the barriers of their disabilities and enable them to complete each year of study at the level of understanding required to ensure college and career readiness. Refocusing preservice and in-service training of teachers so that they have the skills to recognize these varied pathways students can take to proficiency is an essential complementary priority as new innovative systems of assessment are developed. If teachers do not know how to accelerate learning of students who are in the achievement gap, that gap will never be able to be closed. No assessment innovation will close the achievement gap unless teachers know how to close gaps in instruction and access to the curriculum.

The use of accommodations during formative, interim, and summative assessments should be aligned. Yet, this is not the case. Some commonly accepted accommodations for instruction and summative assessments, such as providing a Braille version for blind students, are treated by several formative assessments as yielding invalid results. Without thoughtful alignment across assessments, it is likely that there will be systematic exclusion of students with disabilities from the benefits of formative and interim measures.

Innovation assumes that old wine is not re-bottled in new skins. Current adaptive testing models generally assume that statistical item difficulty reflects the path that all students take to competency. For example, this assumes that all students who do not decode or do basic computation also do not know more typically challenging content. New ways of adapting tests do exist, but they are not generally used in K–12 settings. Current adaptive tests tend to route students into pathways that are similar to out-of-level testing pathways, but the pathways are not transparent; because the computer adapts on the fly, it is often hard to tell what the student actually experienced. In paper and pencil out-of-level testing, it is easy to document when a student is taking a lower-level test. An adaptive test could be built to document pathways into the system—where they start and finish, even how they were routed. That is not typically occurring in existing tests. That type of coding of pathways will be essential to understand what test scores mean and if researchers hope to study the validity of use for the purpose of closing the achievement gap for students with disabilities and other student groups as well.

**Common Alternate Assessment Consortia**

The RTTT assessment consortia are focused on the regular assessment. But students with significant cognitive disabilities also participate in each state’s assessment system by taking an alternate assessment based on alternate achievement standards (AA-AAS). Students who participate in this assessment must demonstrate their knowledge and skills on the same grade-level content standards as all other students, but they are held to different achievement standards, which are different expectations for how they interact with that content and how well they perform on it.10

The Office of Special Education Programs in the U.S. Department of Education has provided funding for consortia of states to develop common AA-AAS based on the Common Core Standards. These consortia will be developing participation criteria for the AA-AAS as well as defining the content to be assessed. It is essential that the RTTT assessment consortia coordinate with the Common AA-AAS Consortia so that there are no gaps in who is included in the assessment system via the two types of assessments and so that the assessed Common Core Standards are coordinated across the consortia. To do this, there will need to be regular communication between the consortia.
An Assessment System That Is Better for All Students

Carefully thinking about all students in the assessment population, including those who receive special education services and those special education students with significant cognitive disabilities, can create a better educational system and better educational assessments for all students. This system must be broad in conception, and it must focus on instruction and professional development for teachers even as it focuses on the assessment system.

All groups of students need to be visible in the results. It is also important to ensure that the assessments developed by the RTTT assessment consortia not have more negative implications for special education students than for other students. For example, special education students generally are more mobile and have poorer attendance than other students. Some models might exclude students with missing scores—or create an estimated score. If special education students have more missing scores than other groups, the estimated scores will be less reliable. Similarly, a high-quality assessment system should not mask the scores of any group of students by enabling the progress of one group to compensate for poor performance of another group.

Schools and districts where students with disabilities are achieving at higher levels show that special education students can achieve the same outcomes as other students when certain conditions are in place. The characteristics of schools and districts that are highly effective for special education students are consistent: strong leadership, a focus on the content and achievement standards, inclusion of all students in grade-level instruction, good data-driven approaches to instruction, and a system-wide belief that the school is responsible for all students and that there are no excuses for poor student performance.
Several specific actions are recommended here for the RTTT assessment consortia to meet the needs of special education students.

1. **Develop a set of common accommodation policies for the RTTT assessments.**

   The RTTT assessment consortia should create a set of accommodations policies for the common assessments they develop. These policies will need to cover all assessments—interim, formative, summative, and others—that are considered part of the assessment system. Because states currently have different accommodation policies and perspectives on accommodations, it will be important for the RTTT assessment consortia to use a process that recognizes these different backgrounds and perspectives as a set of common accommodations are developed.

   Several steps are recommended for the consortia (see Appendix A for details). These include starting from precise definitions of the content to be assessed. It is also important to identify what is not intended to be assessed but may pose a challenge for students with certain disability characteristics. Expert stakeholders should be convened to clarify the nature of the test design and any additional changes that might be needed to accommodate special education students. This group should also determine what professional development is needed to ensure understanding and monitor implementation of the developed common accommodations policies.

2. **Follow accessibility principles in development, field testing, and implementation.**

   Accessibility, also called universal design, refers to a wide range of approaches to making a test appropriate for all students, including special education students. In addition to considerations for the design of the test, accessibility includes ensuring that special education students are included when items are field-tested and that the test results of special education students are considered when items are reviewed.
and flagged. It also includes removing items that display disability bias. To ensure that accessibility is considered throughout the development and implementation process, the RTTT assessment consortia should engage stakeholders to work with others on the test design (including working with the accommodations policies group) and set up procedures to ensure that special education students are included in every phase of the process: development, pilot testing, field testing, and implementation.

3. **Ensure that the design of computer-based tests is appropriate for special education students as well as for other students.**

It will be important for the RTTT assessment consortia to agree on features of the computer-based test that are available to all students (such as increased text size) versus those that are available only to special education students and Section 504 students, or to English language learners, or to any student for whom an educator has identified a need. Thinking through a variety of considerations for computer-based testing will be critical (see Appendix B).

Just as it is important to have expert stakeholders focused specifically on accommodations policies and on accessibility, it will be important to have these stakeholders focused on the design and accommodations for a computer-based test. After determining which features are available to all students, test developers need to consider accessibility in conjunction with procedures for meeting item interoperability and portability across technology platforms (see Appendix C for a white paper addressing this topic).

4. **Develop formative and interim assessments to ensure inclusion of special education students in grade-level curricula focused on accelerated learning.**

The RTTT assessment consortia should use the same processes for their formative and interim assessments as are recommended for addressing inclusion, accommodations, and accessibility in their summative assessments. The goal is to ensure that these assessments assist educators in accelerating the learning of special education students rather than relegating these students to remediation on basic skills without access to important grade-level content. The focus on formative and interim assessments, in particular, will require consideration of learning progressions so that educators better identify which skills are essential to catching up and which are not. In addition, it will require professional development to assist educators in understanding varied content pathways that can help ensure that the learning of all students, especially special education students, is focused on essential skills and knowledge for college and career readiness.

It is particularly important that careful consideration be given to adaptive testing models used for formative and interim assessments, to ensure that they not reinforce lower-grade-level learning. The purposes of each test should be clearly articulated. The use of computerized adaptive testing or vertical scales for the educational purpose of closing the achievement gap requires evidence from test developers and users to demonstrate that such test administration designs and scaling models help achieve the testing purpose. Each form of testing needs to be defended with a validity argument that ensures fairness for each subgroup of students, including special education students. The pathways that students follow through the test should be documented so that the validity of this approach for students who are more likely to follow different learning progressions can be examined, all with the goal of ensuring that the assessment systems of the RTTT assessment consortia help close the achievement gap rather than exacerbate it.
5. Communicate and coordinate with the alternate assessment consortia.

To have a truly inclusive assessment system, all aspects of the system must be coordinated and cohesive. Because a portion of students in schools today (those with significant cognitive disabilities) are held to different achievement standards and participate via the alternate assessment based on alternate achievement standards (AA-AAS), the regular assessment and the AA-AAS must be coordinated and form a comprehensive assessment system. Failure to do so from the beginning will likely leave gaps in the assessment system because it will not be able to incorporate all students.

To form a strong assessment system, the RTTT assessment consortia and the AA-AAS consortia must have conversations with each other about each aspect of the system. Content is an essential part of this discussion, as are the participation criteria for the regular assessment and the AA-AAS. A cross-consortia partnership should be formed across the two types of funded consortia (regular and AA-AAS) to ensure that these conversations take place.
Careful planning will help ensure that the RTTT assessment consortia move forward in a way that builds on all that is known rather than haphazardly moving in a way that may make tests less accessible and less valid for some students with disabilities.

Final Note

The recommendations that have been proposed here for meeting the needs of special education students all start with communication about accommodations, accessibility, and the development process for all students, as well as involvement of expert stakeholders. These seemingly separate stakeholder groups could be a single group that cuts across all topics related to inclusive assessment systems, a strong nationally representative accessibility/inclusion expert panel similar to and complementary with a technical advisory committee. The individuals participating in these groups or a single group should understand grade-level content, special education students, and assessments.

It is important to realize that the research base for developing assessments that are more appropriate for all students, including special education students, has dramatically increased over the past several years. Based on this research, the National Center on Educational Outcomes (NCEO) developed six principles for assessments used for accountability (see Appendix D). Each of these principles is supported by specific characteristics of assessment systems that are appropriate for all students, including special education students. The National Accessible Reading Assessment Projects (NARAP) also developed a set of principles to guide the development and implementation of reading assessments (see Appendix E). Each of these principles is supported by specific implementation guidelines and an annotated bibliography.

The NCEO and NARAP principles are an essential component of judging the quality of innovative assessment systems and are a necessary complement to the revised *Standards for Educational and Psychological Testing*. They reflect the whirlwind of research and development activity to support better assessments for every student, including special education students. They also provide a foundation for actions that the RTTT assessment consortia should take to meet the needs of special education students.

Careful planning will help ensure that the RTTT assessment consortia move forward in a way that builds on all that is known
rather than haphazardly moving in a way that may make tests less accessible and less valid for some students with disabilities. The time is right to build innovative assessments that are best for all students rather than for a subset of the students in U.S. schools today.
1. **Define precisely the content to be assessed and what is not the focus of assessment.** It is essential that during test conceptualization the constructs to be measured are precisely and explicitly defined. This effort should be completed with input from content experts as well as assessment experts. It should clarify which knowledge and skills are the target of the assessment and which are not.

2. **Identify the specific task requirements of each assessment, focusing on which elements of the tasks are critical to the content being assessed and which are not.** For each type of assessment task, whether it is a multiple-choice or constructed response item on a computer, or a performance task in the classroom, the specific elements of the task requirements that are essential to the content for the assessment should be listed. Similarly, those that are not essential to the content and could be addressed through the test design or accommodations should be listed as well.

3. **Discuss the characteristics of special education students and how these characteristics may affect the content to be assessed and the tasks that will be used.** It may be easier to start with characteristics associated with sensory characteristics and then move to those that involve behavior or learning. Lists of these types of characteristics (such as mobility limitations, short-term memory difficulties, attention difficulties, anxiety conditions) have been developed but should be reexamined and clarified through discussions with both general education and special education teachers.

4. **Reach agreement on the characteristics of the test design that are available to all students and determine whether the use of these design features need to be documented.** With the innovative assessment proposed by the RTTT assessment consortia, the opportunity is greater than ever to produce assessments that are as accessible and universally designed as possible. This means that the test can be designed in a way to reduce the need for accommodations. To do this, the test design must eliminate or minimize the effects of extraneous factors such as providing white space and avoiding unnecessary linguistic complexity when that is not the focus of assessment. Item developers must be trained on these concepts so that extraneous barriers are not introduced during item development, and items and assessments should be reviewed using tools that determine the design strengths and weaknesses before field testing. The goal of all these activities is to ensure that the assessment produces valid results for all students.

5. **Reach agreement on whether accommodations are available to all students or some subset of students.** Special education students are not the only students who may require accommodations. Other groups include students with Section 504 accommodation plans and English language learners (including those with disabilities). In addition, after discussing the purpose of accommodations and the continuum of
strengths and weaknesses in the general population, it may be determined that all students should have access to some or all accommodations (in addition to the test design features that should be available to all students).

6. **Reach agreement on the accommodations that may be used by individual students.** At this point, a list should be created of accommodations that produce valid results for each assessment in the assessment system. A mutually agreed upon organization for the lists should be developed to aid those who are to make decisions about the accommodations that individual students will use.

7. **Develop a plan of professional development on the common accommodation policies.** Even the best accommodations policies are worth little without professional development to ensure that those making decisions know the steps involved in the decision-making process, how to determine accommodations needed in the classroom, and how they relate to assessment accommodations. There are several examples of training programs for those who make decisions about accommodations. These programs should be examined and adopted or adapted for use by the RTTT assessment consortia.
APPENDIX B

Computer-based Testing Considerations

1. Consider the assumptions and beliefs of various stakeholders about computer-based instruction and assessments.

2. Consider the system as a whole, from the computer infrastructure to classroom and instructional experiences with computers.

3. Consider the computer or online platform first, with input from individuals who know students with disabilities and their accessibility needs.

4. Consider a process for bringing in the needed expertise to delineate the specific accessibility features of computer-based testing, and to determine what specific accommodations may still be needed by students with disabilities, as well as to determine whether a computer-based test may inadvertently create new accessibility issues.

5. Determine the policies for which accessibility features will be made available to all students and which are designated for specific groups of students, such as students with disabilities.

6. Consider how to track the use of accessibility features incorporated into computer-based testing design.

7. Field-test the accessibility features of the computer-based test at the same time that the computer-based test is field-tested.

8. Examine results from computer-based testing for students with disabilities to determine whether there are any features or characteristics of the assessment that might need reconsideration.

9. Develop training for teachers and students to ensure that students benefit from accessibility features.
APPENDIX C
Meeting Interoperability and Accessibility Requirements by Adopting the Accessible Portable Item Profile (APIP) Standards

A WHITE PAPER PREPARED BY
DIRK MATTSON, MINNESOTA DEPARTMENT OF EDUCATION
MICHAEL RUSSELL, NIMBLE ASSESSMENT SYSTEMS

Opportunity for Interoperability
The Race to the Top Assessment Notice Inviting Applications (NIA) establishes two important requirements that can enhance the quality and efficiency of common assessments. The first of these requirements focuses on “assessing all students, including English learners and students with disabilities” using appropriate accommodations. To accomplish this, applicants must develop appropriate accommodations for common assessments and, importantly, “use technology to the maximum extent appropriate to develop, administer, and score assessments.” This requirement holds potential to improve the validity of test-based inferences for students with disabilities and is best accomplished by applying principles of Universal Design throughout the assessment development process.

The second NIA requirement focuses on “maximiz[ing] the interoperability of assessments across technology platforms and the ability for States to switch their assessments from one technology platform to another by developing all assessment items to an industry-recognized open-licensed interoperability standard.” Prior to this year, the Question and Test Item (QTI) Interoperability Specification came closest to establishing an industry-recognized standard. QTI was developed by the IMS Global Learning Consortium (GLC) which is a global, nonprofit, member organization that provides leadership in shaping and growing the learning industry through community development of interoperability and adoption practice standards. Today, several US and International testing companies are members of the IMS GLC.

The NIA provides an opportunity for consortia to capitalize on work already underway among many states and currently funded by the US Department of Education.

Where We Are Today
The IMS GLC developed QTI in an effort to promote interoperability and portability of tests and test content (e.g., test items) across delivery platforms (e.g., test delivery engines and learning platforms such as Blackboard). In its current form, QTI contains a large corpus of codes (a.k.a., tags) that can be used to specify elements of a test or a test item. As an example, one set of tags can be used to specify an item type (e.g., multiple-choice, short answer, matching, etc.). Another set of tags can be used to specify the correct answer for an item and specific actions that might be taken when a wrong answer is selected (e.g., a specific “hint” is provided when a specific wrong answer option is selected). While QTI contains a comprehensive list of tags that are currently being used internally by many testing companies, there are three shortcomings of QTI with respect to meeting the requirement of an industry-recognized interoperability standard for RttTA common assessments.

First, QTI is currently designed to be a specification that contains a corpus of tags that provide information about tests and test items, but does not establish a clear description of the intended behaviors that are expected to result when a tag is applied. This lack of specificity
allows computer programmers to apply their own interpretations to the behaviors that result when a tag is encountered. In turn, different interpretations result in a lack of standardization in how QTI tags are applied and the resulting behaviors that occur when a QTI tag is encountered by a test delivery program.

Second, it is necessary to provide instructions on how to make a given item accessible. While QTI currently provides programmers with a comprehensive set of tags that can be used to identify specific elements of a test item, none of these tags are designed to provide instructions on accessibility. A separate IMS GLC initiative, called Access for All, was launched to develop a set of tags that focus on accessibility. Like QTI, Access for All specifications contain a corpus of tags that specify different ways to present content to improve accessibility for different users (e.g., presenting text-based content in Braille for users who are blind and read Braille, presenting verbal descriptions of graphics for users with reduced vision, presenting auditory information in American Sign Language for users with hearing needs, etc.). In its current form, Access for All is independent of QTI and is not viewed as a component of an industry-recognized interoperability standard. As a result, even if specific behaviors were defined for QTI so that it could provide interoperability, the fact that Access for All is not currently embedded with QTI would result in test items that are interoperable, but not necessarily accessible.

Third, the current set of Access for All tags provides specifications for many, but not all, of the accommodations and access needs required by students who are expected to perform a common assessment. In order to provide access for all students, the Access for All tags must be expanded to cover the full domain of expected accommodation and access needs provided in a technology-based test delivery environment. In addition, like QTI, specific behaviors that are expected to result when a given Access for All tag is applied have not been defined. Thus, in order to standardize the behaviors that result when an Access for All tag is applied, behavior specifications must be developed.

**Where We Can Go Tomorrow**

To overcome these shortcomings and to develop a standard that the industry can apply to improve the accessibility and interoperability of test items, the US Department of Education has provided funding for the Accessible Portable Item Profile (APIP) project. Led by the Minnesota Department of Education, the APIP project consists of a consortium of 8 member states and 3 observing states that are working closely with the IMS GLC and Nimble Assessment Systems to define a corpus of tags and associated behaviors that can be used to standardize the interoperability of item content and the accessibility components defined for each item. To accomplish this, the APIP project is employing the current set of QTI and Access for All specifications to develop an integrated set of tags, along with descriptions of expected behaviors, that can be applied to standardize the interoperability and accessibility of test items.

Given the advanced state of the APIP project, the involvement of several states along with leaders in interoperability standards and accessible computer-based test delivery, and growing interest from others within the field of testing, APIP serves as the most powerful, comprehensive, and advanced vehicle for meeting the NIA requirements for adopting an industry-recognized interoperability item format standard that empowers common assessments to be as accessible as possible for the widest spectrum of student needs.

**Fulfilling the Interoperability Requirement of the NIA**

The APIP project is developing open-source, open-license interoperability standards that any assessment program, vendor, or other
organization can freely adopt. If adopted by a consortium, the APIP standards allow the consortium to provide assurances that items and assessment instruments developed with USDE funds will fulfill interoperability requirements. As a USDE federally-funded project (Grant #S368A090010), the APIP standards are intended to define a single standard for accessibility and interoperability of assessment items throughout the country. For a consortium that wishes to employ the APIP standards, the following text may be freely used to state how interoperability requirements will be fulfilled:

“To maximize the interoperability of assessments across technology platforms and allow assessment programs that adopt Common Assessments developed by this consortium to efficiently and cost-effectively change technology-based delivery platforms, we will adopt the Accessible Portable Item Profile (APIP) standards. The development of the APIP standards is supported by USDE funding to establish an industry format standard for highly accessible interoperable test items. The APIP standards incorporate key elements of established Question and Test Item specifications, Access for All specifications, and the National Instructional Materials Accessibility Standards to create a single standard for accessible item file format, accompanied by documentation of intended behaviors when the standardized APIP tagging structure is applied to test items.”
PRINCIPLE 1: All students are included in assessments in ways that hold schools accountable for their learning.

Characteristic 1.1. All students are included in every aspect of assessment for system accountability.

Characteristic 1.2. Stakeholders with expertise and experience in varied student learning characteristics collaborate on all aspects of the assessment system to ensure that all students can show what they know and can do.

Characteristic 1.3. The validity of the system is assured through technically defensible assessments that address the implications of varied student learning characteristics.

PRINCIPLE 2: Assessments allow all students to show their knowledge and skills on the same challenging content.

Characteristic 2.1. All students in all settings who receive special education services are included in their enrolled grade level assessment in some way (e.g., in general, accommodated, or alternate assessment), regardless of the nature of disability or special needs.

Characteristic 2.2. All assessments are designed from the beginning with a focus on accessibility for all students.

Characteristic 2.3. Accommodation policies are informed by the defined construct to be measured, available research findings, and the purpose of the assessment.

Characteristic 2.4. Alternate assessments are used to assess the knowledge and skills of students whose disabilities are a barrier to demonstrating knowledge and skills in general assessments under standard conditions or with allowable accommodations.

PRINCIPLE 3: High quality decision making determines how students participate.

Characteristic 3.1. Decisions about how students participate in the assessment system are based on how students show knowledge and skills within the context of accountability testing—not on the student’s instructional program, disability category, or current level of functioning.

Characteristic 3.2. Assessment accommodations are available to all students, and decisions about accommodations are based on an individual student’s accommodation needs and experiences and what the assessment is designed to measure.

Characteristic 3.3. Clear policies, guidelines, procedures, and training on assessment participation decision making are provided for all decision-making partners.

Characteristic 3.4. The IEP team annually reviews and documents assessment participation and accommodation decisions on an individual student basis for each state and district assessment.

PRINCIPLE 4: Public reporting includes the assessment results of all students.

Characteristic 4.1. All students in all placement settings who receive educational services, regardless of severity of disability, are accounted for in the reporting system.

Characteristic 4.2. The number and percentage of students with disabilities assessed and their aggregatable results are reported near to, as often as, and in ways similar to the reporting for students without disabilities.

Characteristic 4.3. The number and percentage of students not assessed or whose results...
cannot be aggregated are revealed in public reports and explanations are given.

**Characteristic 4.4.** Results from assessments administered in ways that raise policy questions are reported separately so that they can be publicly examined and discussed, as well as aggregated with other results.

**Characteristic 4.5.** Reports are provided to educators, parents, students, policymakers, community members, and the media, with a clear explanation of results and implications.

**PRINCIPLE 5:** Accountability determinations are affected in the same way by all students.

**Characteristic 5.1.** Performance data for all students factor into accountability determinations regardless of how they were assessed or why they were not assessed.

**Characteristic 5.2.** Accountability plans treat all groups of students the same.

**Characteristic 5.3.** Formal and informal accountability reports focus on identifying areas and activities that can be changed to improve student learning rather than implying that low performance is attributable to student characteristics.

**PRINCIPLE 6:** Continuous improvement, monitoring, and training ensure the quality of the overall system.

**Characteristic 6.1.** The quality, implementation, and consequences of student participation decisions are monitored and analyzed, and the data are used to evaluate and improve the quality of the assessment process at the school, district, and state levels.

**Characteristic 6.2.** States provide training to multiple stakeholders about the implications of use of available assessment options to improve IEP team decision making about how the student participates in the large scale assessment for accountability.

**Characteristic 6.3.** The use that is made of accountability reports and the impact that accountability decisions have on educational processes and student learning are monitored to determine what adjustments may be needed to improve the accountability system.

**Characteristic 6.4.** The quality of assessment tools is continuously evaluated and improved by applying information gathered about the use and impact of assessment results and by responding to developments in the field of measurement.
APPENDIX E

NARAP’s Principles for Accessible Reading Assessments

PRINCIPLE 1: Reading assessments are accessible to all students in the testing population, including students with disabilities.

Guideline 1-A. Understand and account for the range of student characteristics and experiences that impact reading when designing reading assessments.

Guideline 1-B. Begin the development of reading assessments by applying the elements of universal design.

Guideline 1-C. Use technologies and other evidence-based approaches to provide all students with a variety of assessment options within a similar testing experience.

Guideline 1-D. Document decisions that are made to make tests more accessible, and monitor the effects for different groups of students.

PRINCIPLE 2: Reading assessments are grounded in a definition of reading that is composed of clearly specified constructs, informed by scholarship, supported by empirical evidence, and attuned to accessibility concerns.

Guideline 2-A. State standards are grounded in an evidence-based definition of reading.

Guideline 2-B. Design reading tests to allow for individual measurement of the components reflected in state standards.

Guideline 2-C. Use criteria to select texts that represent different genres and promote the use of interesting passages that are likely to engage all students.

Guideline 2-D. Use criteria to determine the visual elements that should be included within texts while avoiding distracting elements that impact students’ comprehension.

Guideline 2-E. Present reading tasks that students perceive as worthwhile and that enable them to be self-efficacious as learners.

Guideline 2-F. Ensure that test blueprints are aligned with the state standards.

PRINCIPLE 3: Reading assessments are developed with accessibility as a goal throughout rigorous and well-documented test design, development, and implementation procedures.

Guideline 3-A. Initial test design considers the characteristics of all test takers.

Guideline 3-B. Item development and evaluation considers the characteristics of all test takers.

Guideline 3-C. Test assembly and evaluation considers the characteristics of all test takers.

Guideline 3-D. Document the steps that have been taken to ensure that the characteristics of all test takers have been considered.

PRINCIPLE 4: Reading assessments reduce the need for accommodations, yet are amenable to accommodations that are needed to make valid inferences about a student’s proficiencies.

Guideline 4-A. Begin the assessment development or revision process by reviewing allowed accommodations to determine whether they could be incorporated into the design of the assessment.

Guideline 4-B. Identify and determine the essential accommodations that are still needed after incorporating as many as possible into the assessment.

Guideline 4-C. Develop a strong rationale and
evidence to support the validity of inferences from assessment results when accommodations are provided.

Guideline 4-D. Provide information and support to ensure that accommodations are implemented in an appropriate manner.

Guideline 4-E. Adjust the reading assessment approach to address the needs of some groups of students that cannot be met by typical test design or accommodation procedures.

PRINCIPLE 5: Reporting of reading assessment results is designed to be transparent to relevant audiences and to encourage valid interpretation and use of these results.

Guideline 5-A. Provide clear and concise score reports that are appropriate for a diverse audience.

Guideline 5-B. Pilot score reports with all relevant groups of score users.

Guideline 5-C. Detailed information about the assessment and score results is available in a document that is accessible to all test takers and score users.

Guideline 5-D. Provide information regarding the precision of reported scores for all relevant groups.
Notes

1) Development of this paper was supported by the U.S. Department of Education, Office of Special Education and Rehabilitative Services, Office of Special Education Programs, through a cooperative agreement (H326G050007) with the University of Minnesota, National Center on Educational Outcomes. Opinions expressed in this paper are those of the authors and do not necessarily reflect those of the U.S. Department of Education or its offices.

2) Pie chart provided to the National Center on Educational Outcomes by the National Center for Learning Disabilities.


7) Altman, Thurlow, and Vang, 23.


10) Rachel Quenemoen, A Brief History of Alternate Assessments Based on Alternate Achievement Standards, Synthesis Report 68 (Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes, 2008).

11) For research and case studies, see Donahue Institute, A Study of MCAS Achievement and Promising Practices in Urban Special Education: Data Analysis and Site Selection Methodology (Amherst, MA: University of Massachusetts, 2004) and Candace Cortiella and Jane Burnette, Challenging Change: How Schools and Districts are Improving the Performance of Special Education Students (New York: National Center for Learning Disabilities, 2009).


13) Martha L. Thurlow et al., Accessibility Principles for Reading Assessments (Minneapolis, MN: National Accessible Reading Assessment Projects, 2009).