



TN

Department of
Education

Tennessee Task Force on Student Testing and Assessment

Final Report

Tennessee Department of Education | September 2015



TABLE OF CONTENTS

Executive Summary	3
Purpose of Assessment in Tennessee	4
Task Force on Student Testing and Assessment	7
Members of the Task Force	7
Structure of Task Force.....	8
History and Current Landscape of Assessment in Tennessee	10
Required National Assessments	10
Required State Assessments	11
District Assessments.....	11
Educator Survey Results	12
Assessment Principles	15
Summative (Annual) Standardized Assessments.....	15
Formative (Interim, Benchmark) Assessments	16
Test Preparation and Logistics	17
Recommendations	18
Further Analysis	20
Next Steps	21
Conclusion	25
Appendices	26

Executive Summary

On March 2, 2015, Commissioner of Education, Dr. Candice McQueen, announced the creation of the Tennessee Task Force on Student Testing and Assessment. The task force was formed as a result of feedback from the field about the amount of testing, quality of testing, and associated test preparation. Specifically, the task force's stated goals were to: (1) identify and study best practices in student assessment, (2) ensure local school districts and the state are appropriately using assessments to improve student achievement, and (3) better inform stakeholders about the state assessment program. In order to do this, the task force set out to accomplish the following:

1. Conduct an environmental scan of assessment usage and practices across the state.
2. Establish principles addressing purposes and goals of state assessments relative to locally chosen/designed assessments (i.e., formative assessments).
3. Define appropriate practices associated with these principles that best support decision making at the state, district, school, and teacher levels.
4. Gain insight on ways to best communicate about TNReady to all stakeholder groups.

During the late spring and summer, the task force of 18 members made up of district leaders, school leaders, teachers, parents, students, the Tennessee General Assembly, the State Board of Education, and the Tennessee Department of Education, along with ex-officio members from the Tennessee Education Association (TEA), Professional Educators of Tennessee (PET), and the Tennessee State Collaborative on Reforming Education (SCORE), convened five meetings to analyze survey results and presentations on the state of assessment in Tennessee. The task force quickly developed interest in three primary areas of focus: (1) summative (or annual) standardized assessments driven by the state, (2) formative (or interim) assessments that are locally driven, and (3) test preparation and logistics. Before developing recommendations, the task force developed guiding principles.

The task force made 16 recommendations with specific emphasis in four areas: (1) culture of transparency, (2) test reduction, (3) postsecondary alignment and readiness, and (4) test scheduling and logistics. Specific highlights of these recommendations include **releasing test items** for students, parents, and educators, **eliminating** the kindergarten and first grade standardized test option, **eliminating** the 8th and 10th grade EXPLORE and PLAN tests, providing **expectations** to districts regarding formative assessment usage and communication, creating additional **input opportunities** for parents, and ensuring **higher education** guidance, validation, and usage of TNReady. The task force concluded the report with additional areas for further analysis, including more work on district grading practices and policies and the usage of screening tools in early grades. The task force also recommended its continued involvement in reviewing new information about the state of assessment during this time of assessment transition.

Purpose of Assessment in Tennessee

Assessment is an integral part of education in Tennessee. Since 1983, Tennessee has used summative tests to provide important information about the collective progress of students in our state. In fact, these assessments ultimately revealed that our definition of proficiency as demonstrated on the statewide tests was not aligned with proficiency expectations on the National Assessment of Education Progress (NAEP) or The Nation's Report Card. In 2007, Tennessee received an "F" from the U.S. Chamber of Commerce for "Truth in Advertising about Student Proficiency" in its *Leaders and Laggards* report. While we were identifying large percentages of our students as proficient on 2005 state math and reading exams, smaller percentages posted proficient scores on the NAEP in 2005. As a result, Tennessee began systematically raising expectations through more rigorous standards and an aligned statewide assessment. Since 2009, Tennessee has made remarkable progress in both raising expectations for learning and moving more students to proficiency while ensuring our students are postsecondary and workforce ready. Compared to 2011, 131,000 more students are on grade level in math, and nearly 60,000 more students are on grade level in science.

TYPES OF ASSESSMENTS

For purposes of this report, the following definitions will be used.

Diagnostic: often given at the beginning of the school year, this assessment allows teachers to know where each student is beginning in their understanding of the subject

Formative (Interim, Benchmark) Assessment: measures student learning throughout the year so educators can determine if students are making progress and how best to adjust instruction; for purposes of this report, formative will refer to interim and benchmark assessments as well

Summative (Annual) Assessment: measures student learning at the end of the semester/year; state-level assessments are summative assessments administered to all Tennessee students.

Tennessee's **annual assessments** provide district and school leaders, teachers, parents, and students specific information about student learning in order to improve the education of all students. More students in various subgroups (i.e., black, Hispanic, and Native American; economically disadvantaged; students with disabilities; and English language learners) have moved to proficiency as a result of using annual tests to highlight achievement gaps and to measure continuous improvement with all students. Results from annual tests assist teachers and parents in understanding if students have met the learning expectations for the year. Additionally, Tennessee's

annual assessments provide feedback to all of the stakeholders who invest in our students to ensure that funds are being used well and that we are setting our students on a pathway to success.

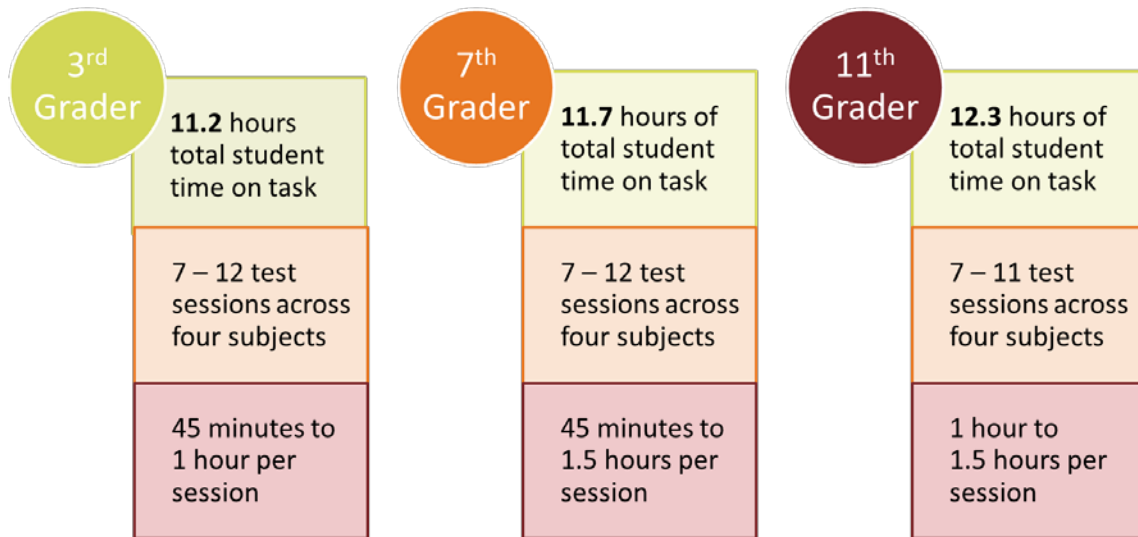
Beyond statewide assessments, Tennessee educators use various additional types of tests to monitor student progress throughout the school year. These **formative assessments** range from teacher-made tests to school-made common assessments to vendor-created benchmark or interim assessments. Typically, formative assessments complement the standards and highlight progress students are making toward annual goals as measured at various points during the school year. Teachers and school leaders primarily use formative tests to help them develop interventions for students who are not making progress or to plan for re-teaching or acceleration of particular standards with groups of students. Educators may also use formative assessment to expose students to samples of state-test questions and the state-test platform or environment. This helps students gain familiarity and ease potential test anxiety in preparation for the summative statewide assessment.

Tennessee educators have been in a period of change during the past several years with a move to new state standards in math and English language arts that center on critical thinking, problem solving, and performance skills—all required by postsecondary and the workforce. Educators have also been preparing for a new assessment that contains test items aligned to the more rigorous state standards. During this time, the department administered several pilot or field tests, such as the Constructed Response Assessment¹, to gather important information for the state’s transition to a new test. Initially, the state planned to give the Partnership for Assessment of Readiness for College and Careers (PARCC) assessment in 2014-15, but when legislation was passed in the spring of 2014 that withdrew the state from the PARCC consortium, the department extended historical Tennessee Comprehensive Assessment Program (TCAP) tests for another school year. Tennessee then led a competitive process to determine a new test provider for 2015-16. As a result, many districts used a variety of vendor-made formative assessments in an effort to gauge preparation for the new, more rigorous annual tests. Some districts administered multiple benchmark tests due to the confusion over PARCC both before and after legislative action.

In recent years, concerns have arisen about the quantity and quality of assessments. Some of this concern was directly connected to the number of formative assessments administered by districts and pilot tests administered by the state during the transition period. Other concerns were related to the amount of test preparation and testing logistics issues leading up to and during annual test administration. As a helpful reference point, the figure below shows how much time a student in the 3rd, 7th, and 11th grade will actually spend on taking the state-required summative assessments for those individual grades in 2015-16.

¹ An assessment that contains test items which are open-ended and require students to “construct” a response without having access to choices or suggestions.

**ACTUAL TIME SPENT TAKING STATE-REQUIRED
SUMMATIVE ASSESSMENTS IN 3RD, 7TH, AND 11TH GRADE**



Tennessee’s new and improved English language arts and math TCAP tests, called TNReady, will inherently change how we design and use formative assessments throughout the year and how we prepare students for TCAP. The transition provides an opportune moment to step back and consider potential improvements to assessment practices across the state.

Task Force on Student Testing and Assessment

In the spring of 2015, Dr. Candice McQueen, commissioner of education, formed the Tennessee Task Force on Student Testing and Assessment. The task force was formed to address concerns about the amount of testing, quality of testing, and associated test preparation. The task force's goals were to: (1) identify and study best practices in student assessment, (2) to ensure local school districts and the state are appropriately using assessments to improve student achievement, and (3) to better inform stakeholders about the state assessment program.

In order to do this, the task force set out to accomplish the following:

1. Conduct an environmental scan of assessment usage and practices across the state.
2. Establish principles addressing purposes and goals of state assessments relative to locally chosen/designed assessments (i.e., formative assessments).
3. Define appropriate practices associated with the principles that best support decision making at the state, district, school, and teacher levels.
4. Gain insight on ways to best communicate about TNReady to all stakeholder groups.

The task force met five times across late spring and summer to learn more about current practices statewide and the local, state, and federal contexts related to assessment. Each meeting included presentations of relevant information and discussion by the task force members about the topics.

This report provides information on the following:

- Task force members, organization, and meeting information
- An overview of the history and current landscape of assessment in Tennessee
- Principles for assessment
- Task force recommendations and areas for further study

Presentations and materials from the task force meetings are included in the appendix.

Members of the Task Force

The task force included representatives of various stakeholder groups including district leaders, school leaders, teachers, parents, students, the Tennessee General Assembly, the State Board of Education, and the Department of Education. The members brought varied experiences and opinions about assessment that led to rich discussion of the topics covered in the meetings. A full list of members is included in the table below.

TASK FORCE MEMBERS

Member	Title
Nancy Ash	Assistant Director of Schools, Lebanon Special Schools
Virginia Babb	Member, Knox County Parent-Teacher Association
Harry Brooks	Chairman, House Education Administration and Planning Committee
Jasmine Carlisle	11 th -grade student, Mt. Juliet High, Wilson County Schools
Phillip Eller	Teacher, Cedar Grove Elementary, Rutherford County Schools
John Forgety	Chairman, House Education Instruction and Programs Committee
Dolores Gresham	Chairman, Senate Education Committee
Bill Harlin	Principal, Nolensville High School, Williamson County Schools
Sara Heyburn	Executive Director, State Board of Education
Valerie Love	Teacher, Dobyns-Bennett High, Kingsport City Schools
Rebecca McBride	Teacher, Brighton High, Tipton County Schools
Sharon McNary	Principal, Richland Elementary, Shelby County Schools
Candice McQueen	Tennessee Commissioner of Education
Mary Reel	Director of Schools, Milan Special Schools
Debbie Shedden	President-Elect, Tennessee School Boards Association; Board Member, Hawkins County Board of Education
Wanda Shelton	Director of Schools, Lincoln County
Beth Unfried	Director of Elementary Schools, Clarksville-Montgomery County Schools
Mike Winstead	Director of Schools, Maryville City

The task force members were supported in their work by ex officio members from the department and other partner organizations including SCORE, TEA, and PET.

Structure of Task Force

The task force met in Nashville for four half-day meetings and one additional full-day meeting between April and August 2015. Each meeting included presentations designed to provide the task force with information about various issues related to assessment. There were both large and small group discussions about the topics presented and the principles and recommendations that were ultimately developed by the group. The general topics discussed at each meeting are outlined below.

SCHEDULE AND CONTENT OF TASK FORCE MEETINGS

April 6, 2015

- Overview and purpose
- Assessment purposes, definitions, state requirements
- Landscape of assessment in Tennessee
- Feedback on SCORE district survey

May 20, 2015

- District, principal and teacher survey results (SCORE)
- Student assessment outcomes in Tennessee

June 15, 2015

- Discussion of other state reports
- Quick scores and test design overview
- ACT
- K-2 assessment

July 15, 2015

- 2015 TCAP results
- Focus group results on assessment from SCORE
- 2015 educator survey assessment results
- Accountability waiver application
- Discussion of draft principles and recommendations

August 24, 2015

- Review of final task force report including principles and recommendations

The agendas and presented materials from each meeting can be found in the appendix. The task force also received information on stakeholder perspectives on assessment from multiple surveys conducted by the department as well as surveys, focus groups, and interviews conducted by the SCORE.

History and Current Landscape of Assessment in Tennessee

Tennessee has a long history of assessments in education. Statewide assessment in Tennessee began with the Tennessee Proficiency Test in 1983. In 1988, the State Board of Education commissioned the Tennessee Comprehensive Assessment Act (TCAP), and in 1992, the Education Improvement Act (EIA) made TCAP a state mandated assessment. Today, the assessment landscape is shaped by many entities: federal law, state law, district policy, and school policy.

Required National Assessments

In accordance with §T.C.A. 49-6-6001(b), 11th graders participate in the ACT or SAT assessment, which is used to measure college readiness and determine HOPE scholarship eligibility. This statute also requires an examination be administered in both 8th and 10th grade to provide information on student preparedness for postsecondary success. EXPLORE (grade 8) and PLAN (grade 10), part of the ACT suite of assessments, have been used to fulfill this requirement in recent years.

A sample of Tennessee students in 4th and 8th grade participate in the National Assessment of Educational Progress (NAEP) every two years as required by the No Child Left Behind Act of 2001 and National Assessment of Educational Progress Authorization Act (Public Law 107-279 III, Section 303). NAEP measures Tennessee student achievement compared to the achievement of students in other states. Because all 50 states administer the NAEP exam, it allows Tennessee to compare the educational outcomes of our students against students in other states and determine if our proficiency levels are on par with the rest of the country.

NATIONAL AND STATE ASSESSMENTS TAKEN IN TENNESSEE

Required National Assessments	Required State Assessments	Optional State Assessments
<ul style="list-style-type: none">•EXPLORE (8th grade)•PLAN (10th grade)•ACT (11th grade)•NAEP (student sample)	<ul style="list-style-type: none">•ELA & Writing (3-8)•Math (3-8)•Science (3-8)•Social Studies (3-8)•HS English I, II, III & Writing•HS Algebra I and II•HS Biology•HS Chemistry•HS U.S. History•RTI² Universal Screeners (diagnostic)	<ul style="list-style-type: none">•SAT-10 (phasing out)

Required State Assessments

TCAP is the statewide assessment system and includes assessments required by both federal and state laws. It includes two components: Achievement and End of Course (EOC). Achievement tests are administered annually in grades 3-8 covering English language arts, writing, mathematics, science, and social studies. End of Course (EOC) exams are given at the conclusion of high school courses in English (I, II, and III), Algebra (I and II), biology, chemistry, and U.S. history.

In the 2015-16 school year, TNReady will replace the previous mathematics, English language arts, and writing assessments in grades 3-8 and high school. TNReady will more accurately assess higher-level thinking that is now embedded in Tennessee state standards.

The requirement that districts must implement Response to Instruction and Intervention (RTI²) has resulted in districts adopting universal screeners and progress monitoring assessments that best meet the needs of their district. A universal screener is a periodic short assessment given to all students in order to identify students who need additional interventions. Universal screeners are not assessments in the traditional sense. They are meant to be quick, informative, and non-intrusive and are given three times a year to identify at-risk students. In most scenarios, reading fluency is assessed for one minute with each student while math computation is administered in a group setting and takes five to eight minutes on average amounting to 30 minutes or less annually for all students. Additional brief progress monitoring assessments for students receiving intervention instruction is necessary to regularly assess progress on identified areas of need.

District Assessments

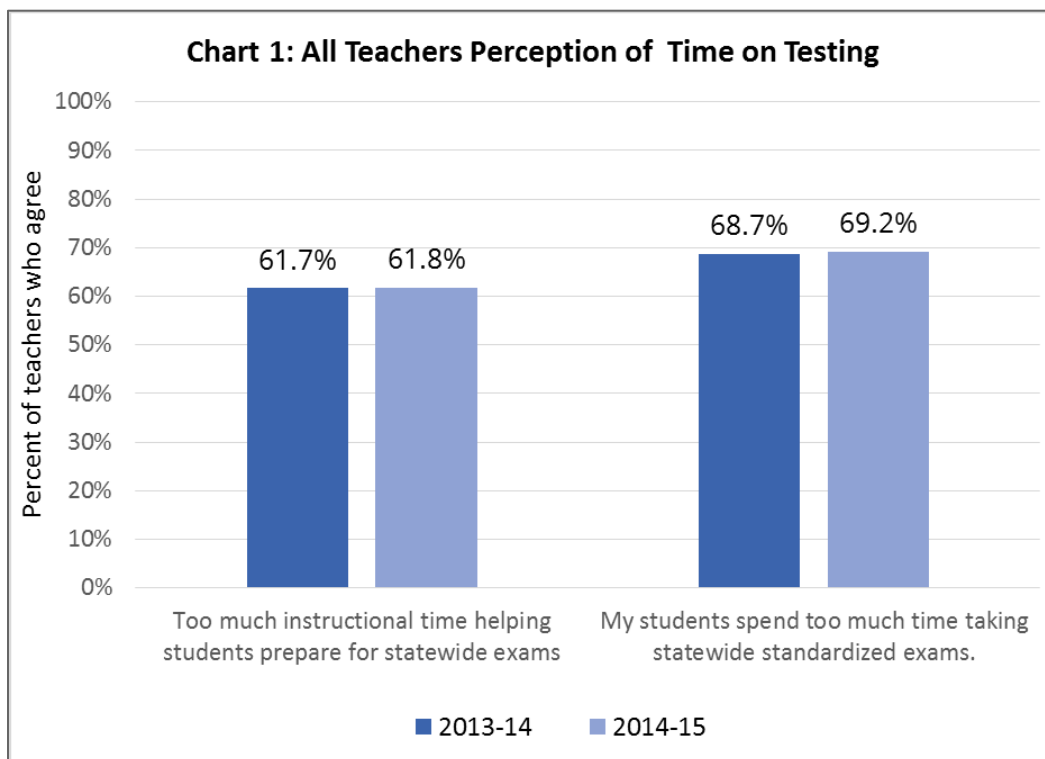
In addition to state-required TCAP and national assessments, students also participate in an assortment of district assessments, which vary in both the number of assessments and the time spent on administration in each district. An assessment that is widely used by the majority of districts is the SAT-10 assessment for grades K-2. The SAT-10 is an assessment **option** for grades K-2 that districts may choose to administer, and it is currently paid for by the state and supported by the Tennessee Department of Education. This assessment provides timely information to teachers, schools, and districts on student learning and helps identify students in need of early intervention and remediation in literacy and numeracy. Results from this assessment may also be used to create individual growth measures for teachers to be used in teacher evaluation.

In addition to SAT-10, districts employ a variety of quarterly benchmark or interim assessments to measure student mastery of specific standards as they track student progress in anticipation of annual statewide assessments. Practices on the frequency and length of these types of assessments vary district by district.

Educator Survey Results

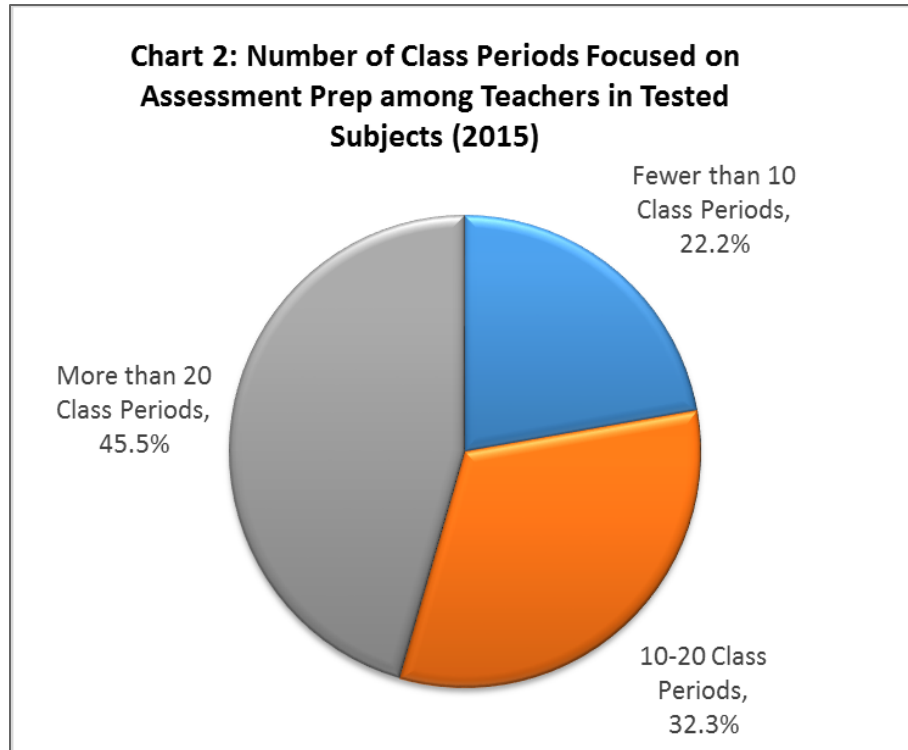
Both SCORE and the department conducted surveys of education stakeholders including district administrators, school leaders, and teachers in the spring of 2015. The analysis from these surveys was shared with the task force to provide a robust picture of educator perception about assessment.

In spring of 2015, more than 36,500 teachers (57 percent) completed the department's educator survey to share their views on a variety of topics. According to this year's survey, educators report overall satisfaction with many areas of their work. However, the subject of assessment does identify challenges. The majority of teachers feel that they are spending too much time on exam preparation and testing. This has remained about the same as in 2014, as shown in Chart 1 below.

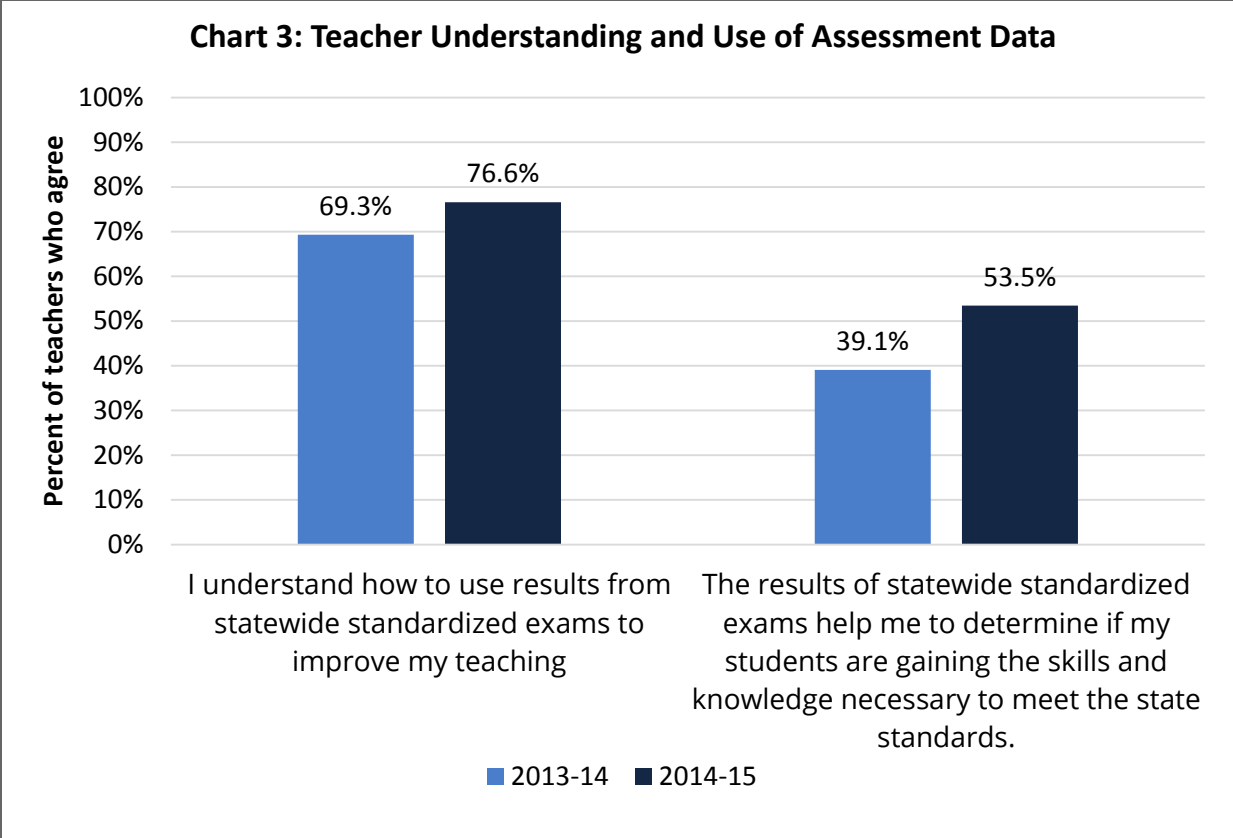


As expected, the level of agreement on this varies slightly among teachers in tested subject areas and teachers in subject areas without a statewide assessment. Nearly three quarters of teachers in tested subject areas report that they spend too much time preparing for statewide exams and that their students spend too much time taking statewide exams.

Nearly half of teachers in tested subject areas report that their students are spending more than 20 class periods on assessment preparation as shown in Chart 2 below.



Although teachers report concerns with the number of assessments and the time it takes to administer them, teachers and other stakeholders value the information provided by the assessments and use it to inform decisions. As compared to 2014, more teachers report understanding how to use data from assessments to monitor student progress and improve teaching. In fact, from last year to this year, there was roughly a 14 percentage point increase in the number of teachers who believe the results from statewide assessments help them understand if their students are gaining knowledge of the state standards. More information about the [2015 Educator Survey](#) is available on the department's website.



The results of the SCORE Assessment Study were similar to those found in the department’s educator survey. More information about the results of the SCORE Assessment Study, which included surveys, focus groups, and interviews, can be found in the [SCORE report](#). SCORE shared the findings from this work with the task force to inform its understanding of the current landscape when developing principles and recommendations.

Assessment Principles

In combination with their prior knowledge, expertise, and experience, the task force members applied what they learned through the task force meetings to develop principles for assessment. These principles, which focus on summative assessments, district benchmarks, and test preparation and logistics, are meant to inform decisions about assessments at the state, district, and local level. The principles address the benefits of the assessment and identify characteristics of each area of focus.

Summative (Annual) Standardized Assessments

Tennessee students benefit from **summative (annual) standardized tests** for the following reasons:

1. Test results and related information show progress toward mastery of standards.
2. Test results and related information indicate post-secondary readiness.
3. Experience taking annual tests allows students to demonstrate cumulative knowledge and retention of this knowledge.
4. Test results enable schools to measure progress and identify the needs of each child to provide feedback to parents and to identify next steps in student learning for educators.
5. Experience taking annual tests prepares students for college and career entrance exams, such as the ACT and SAT, and provides families with information about readiness for postsecondary.

Tennessee's **summative standardized assessments** should:

1. Be properly aligned to the state's standards.
2. Embody the full range of expectations in the state's standards and demonstrate the level of rigor needed in the process of teaching and learning.
3. Ensure equitable access to the state's standards. Specifically, disaggregated test information should help educators know if they are serving all student subgroups well and maintaining universally high expectations.
4. Show *all* students' culmination of understanding, strengths, and areas for improvement to stakeholders.
5. Serve as a benchmark for Tennessee against other states and countries.
6. Be addressed in IEPs in the form of guidance for educators in determining participation and accommodations for students served by IEPs.
7. Be part of the instructional process and the cycle of learning.
8. Be used for school improvement planning, to inform readiness for instruction, and support students. Standardized tests should not be the only source of information for student progress monitoring or daily instructional planning.
9. Allow for transparency to the public on question types, subjects, and reporting by standard/cluster level while still maintaining test integrity and budget efficiency.
10. Have associated reports that are clear and readily understood by educators, students, and parents, and provide timely feedback.

Formative (Interim, Benchmark) Assessments

Tennessee students benefit from **formative assessments** for the following reasons:

1. Formative assessments indicate individual student “standing” and progress toward both classroom-level goals and annual grade-level standards for learning.
2. Formative assessments serve as teaching tools that clarify for students what is most important in a curriculum.
3. Formative assessments provide information that leads to daily decision-making such as re-teaching, intervention, or enrichment needs. Formative assessments should guide instruction.
4. Formative assessments, in the form of benchmark tests, indicate school-level progress toward mastery of standards.
5. Formative assessments facilitate student ownership of their learning.

Formative assessments should:

1. Provide timely feedback to teachers, parents, and students.
2. Be used primarily to 1) measure mastery on standards that have been taught in alignment with the expectations of the standards and the rigor of the summative assessment, and 2) provide diagnostic information about student skill gaps and enrichment opportunities.
3. Be used for specific decision-making and next steps so teachers and students can see progress toward mastering standards and building skills.
4. Be carefully selected or designed and lead to defined student outcomes.
5. Be teacher-led as much as possible because teacher-created and school/district-created common assessments best inform instructional changes. These customized assessments are preferred over generic, large-scale, off-the-shelf assessments that may not be aligned to Tennessee standards.
6. Be addressed in Individualized Education Programs (IEPs) in the form of guidance for educators in determining participation and accommodations for students served by IEPs.
7. Continue to be a district option for large-scale interim or benchmarking assessments. Districts should have flexibility to design their own tests but should have support (e.g., item banks, professional development, etc.) when making tests.
8. Have a clearly communicated purpose that all stakeholders, especially parents, understand.
9. Be designed in a variety of forms and include assessing students on areas that cannot be evaluated with a traditional paper-and-pencil test. Intentional cross-curricular and multi-purpose formative assessments are encouraged.

Test Preparation and Logistics

The **State of Tennessee** should:

1. Provide clear communication and guidance on reducing time spent on specific test preparation.
2. Emphasize and support the idea that the best test preparation is focusing on great teaching and engaged student learning every day.
3. Provide guidance and maximum flexibility during the testing window² to ensure testing logistics issues are minimized and that the entire school year continues to be utilized for quality instruction even during the testing cycle.
4. Ensure transparency in the test design – including blue prints and question format – to help teachers and students know what is expected of them in reaching mastery of the standards.
5. Provide more test scheduling and logistics support with opportunities for input from those on receiving end of tests (teachers and students).
6. Continue to assist in greater use of technology for personalized learning and to prepare schools for online testing.

Tennessee school districts should:

1. Strive to continue instruction and avoid “shutting down” during school testing windows.
2. Work toward daily technology usage for every student. Technology must be integrated in the teaching and learning cycle on a regular basis so students can build technology skills.
3. Ensure that students have the opportunity to develop familiarity with the test platform in a natural way and during instruction.
4. Take great care in not passing on stress or test anxiety to students. Where possible, testing should be coordinated by someone other than the school counselor so they are available to support teachers and students experiencing test anxiety.

Based on the principles on summative assessments, formative assessments, and test prep and logistics, the task force also developed a list of 16 recommendations of actions to be taken by the state related to assessments that are listed below.

² The testing window is the window of dates in which districts can choose to administer the TCAP; students only spend a portion of the testing window actually taking the test.

Recommendations

Culture of Transparency

1. The department and districts should continue to focus on **improving communication** around testing and accountability to create clarity, transparency, and trust.
2. Beginning with the 2016-17 school year, the department should annually release as many **summative test items** as possible without compromising test security and development. These should include operational, non-linking test items.
3. The department should annually release **standardized test blueprints**, test specifications, and the methodology for calculating all score reports.
4. The department and individual districts should build upon current reporting requirements related to mandated assessments and clearly communicate to the public the **purpose** of large-scale formative or summative assessment usage. If either the department or an individual district administers a large-scale assessment, teachers, parents, students, and other stakeholders need to know "why." This information should be easily accessible to the public through district and school websites as well as other sources.
5. The department should communicate with educators through regular channels including regional assessment meetings, about **successful practices for test preparation** for annual tests as well as practices that should be avoided. Communication should center on the idea that the best test preparation is focusing on great teaching and engaged student learning every day.

Test Reduction

6. The department should **eliminate** the kindergarten and first-grade annual standardized test option when SAT-10 is discontinued at the end of the 2015-16 school year. The department should create a new second-grade assessment aligned to Tennessee state standards as an alternate to the second-grade SAT-10 and provide this **option** to districts.
7. The department should address the issue of over-testing and possible test redundancy by **eliminating** the mandatory EXPLORE (8th grade) and PLAN (10th grade) tests, and **not adopt** ACT's new alternative ASPIRE. The department should continue to require ACT for all 11th grade students except for the rare circumstances in which an IEP precludes a student from taking the ACT.
8. Districts should carefully select and design formative assessments that lead to defined student outcomes. Districts should **reduce** formative assessments that do not guide decision-making and next steps in instruction. Districts should utilize the principles of formative assessment as presented in the assessment task force report.

9. The department should form a **parent advisory group** to specifically give feedback on concerns related to over-testing and test preparation practices as well as feedback on information desired for annual student test reports.

Postsecondary Alignment and Readiness

10. The department should partner with leadership from **higher education**—University of Tennessee (UT), Tennessee Board of Regents (TBR), and Tennessee Independent Colleges and Universities Association (TICUA)—to validate TNReady assessment as a tool for determining remedial placement, as well as helping to validate TNReady as a measure of progress toward postsecondary and career readiness.
11. The department should expand ACT or SAT **retake** opportunities for all students.
12. The department should ensure annual tests provide **clear reports** for educators, students, and parents that point to alignment to postsecondary readiness.

Test Scheduling and Logistics

13. The department should convene a **testing scheduling and logistics advisory group** that is representative of district personnel across the state to address logistical challenges. This team should focus on test scheduling that minimizes disruptions and provide ongoing guidance and support (in the form of sample schedules and process guides) particularly during the first year of TNReady.

Other

14. The department should create **additional portfolio assessment options** for teachers in non-tested grades and subjects, specifically first grade, from which districts can choose.
15. The department must ensure future and new teachers have explicit knowledge about assessment for learning. Tennessee **teacher preparation programs** should include a specific curriculum or module on assessment with specifics to Tennessee. All new teachers, whether in tested subjects or not, also need specific training with embedded professional learning around this area of the TEAM rubric.
16. The department should work directly with districts to increase awareness of the realities of **test anxiety** while providing specific guidance in how to help educators avoid passing on stress or test anxiety to students. School counselors must be available to assist in this work during test administration.

Further Analysis

There were several areas identified by the task force that require more information in order to better understand the topic or issue. Areas for further research and analysis that the department should explore are outlined below:

1. The department should look more closely at grading policies and practices with districts in an effort to determine whether greater consistency and clarity needs to be addressed.
2. The department should explore the usage and understanding of screening versus testing in grades K-2 in an effort to bring clarity of purpose and practice.
3. The department should review the possibility of purchasing more item banks for district-created and aligned formative assessments.
4. The department should consider conducting trainings on how to create common assessments and how these trainings can be used as learning opportunities for teachers.
5. The department should continue to analyze the concern regarding over-testing and determine strategies to ensure that schools and districts have local assessments that are aligned to standards, yet not duplicative of state standardized assessments.

Next Steps

Each of the 16 recommendations requires immediate action and concrete ownership to ensure they are completed. The following section is an overview of each recommendation and accompanying next steps.

Culture of Transparency

1. The department and districts should continue to focus on **improving communication** around testing and accountability to create clarity, transparency, and trust.
 - Communicate TNReady to media and general public audiences
 - Collaborate on communication, training, etc. with all education stakeholders and partners
 - Create proactive and clear communications to educators
 - Post additional information about the value-added model and quick scores on the department website
 - Require districts to share test results/reports directly with students as well as parents
2. Beginning with the 2016-17 school year, the department should annually release as many **summative test items** as possible without compromising test security and development. These should include operational, non-linking test items.
 - Determine cost
 - Work with appropriate stakeholders on budgetary impact
 - Create plan for public release
3. The department should annually release **standardized test blueprints**, test specifications, and the methodology for calculating all score reports.
 - Have already released these items for 2015-16 school year
 - Prepare clear timelines for release in future years
 - Staff appropriately
4. The department and individual districts should build upon current reporting requirements related to mandated assessments and clearly communicate to the public the **purpose** of large-scale formative or summative assessment usage. If either the department or an individual district administers a large-scale assessment, teachers, parents, students, and other stakeholders need to know "why." This information should be easily accessible to the public through district and school websites as well as other sources.
 - Already posted information on state annual tests on the department's website (also created educator and parent guidebooks on TNReady)
 - Present principles/recommendations of assessment task force to directors at the Superintendent Study Council Conference during September 2015
 - Determine method to ensure compliance

5. The department should communicate with educators through regular channels including regional assessment meetings, about **successful practices for test preparation** for annual tests as well as practices that should be avoided. Communication should center on the idea that the best test preparation is focusing on great teaching and engaged student learning every day.
 - Already scheduled regional assessment meetings for educators during September
 - Already scheduled additional regional TNReady information lunch and learn meetings for educators, parents, legislators, and community members during Sept. and Oct.
 - Create additional messages, guidance documents, and communication tools for district superintendents and educators about principles of assessment

Test Reduction

6. The department should **eliminate** the kindergarten and first-grade annual standardized test option when SAT-10 is discontinued at the end of the 2015-16 school year. The department should create a new second-grade assessment aligned to Tennessee state standards as an alternative to the second-grade SAT-10 and provide this **option** to districts.
 - Communicate elimination of kindergarten and first-grade annual standardized assessment option
 - Draft Request for Proposals to begin process for second-grade assessment
7. The department should address the issue of over-testing and possible test redundancy by **eliminating** the mandatory EXPLORE (8th grade) and PLAN (10th grade) tests, and **not adopt** ACT's new alternative ASPIRE. The department should continue to require ACT for all 11th grade students except for the rare circumstances in which an IEP precludes a student from taking the ACT.
 - Communicate elimination of 8th and 10th grade ACT tests
 - Identify or create career inventory (currently a part of 8th grade EXPLORE test) to be used with 8th grade students
 - Continue to ensure that state standards align to ACT and create TNReady alignment based on postsecondary readiness standards
 - Map 2015-16 Explore and Plan test results to 2015-16 TNReady results
 - Communicate TNReady benchmarks at 8th and 10th grades to ensure readiness to meet benchmarks on ACT at 11th grade
8. Districts should carefully select and design formative assessments that lead to defined student outcomes. Districts should **reduce** the number of formative assessments that do not guide decision making and next steps in instruction. Districts should utilize the principles of formative assessment as presented in the assessment task force report.
 - Communicate expectations of assessment principles to directors at the Superintendent Study Council Conference in September 2015
 - Assist in training districts on effective formative assessments
 - Ensure understanding of and compliance with recommendation four

9. The department should form a **parent advisory group** to specifically give feedback on concerns related to over-testing and test preparation practices as well as feedback on information desired for annual student test reports.
 - Start collecting nominations for parent advisory group
 - Plan first meeting in October
 - Share report options with various parent focus groups as well as parent advisory group

Postsecondary Alignment and Readiness

10. The department should partner with leadership from **higher education**—UT, TBR, and TICUA—to validate TNReady assessment as a tool for determining remedial placement, as well as helping to validate TNReady as a measure of progress toward postsecondary and career readiness.
 - Already met with UT, TBR, and TICUA leadership to review TNReady focus and design
 - Identify UT, TBR, and TICUA math and English and language arts faculty for TNReady item review and validation during fall 2015
 - Include Tennessee higher education faculty in all standards setting processes and decisions during 2016
 - Partner with Tennessee higher education administration and faculty to set TNReady scores for admission into postsecondary programs by December 2016 for new students entering postsecondary during the 2017-18 school year
11. The department should expand ACT or SAT **retake** opportunities for all students.
 - Identify funding to facilitate ACT or SAT test retakes
 - Assist districts in providing additional support to students who do not meet ACT or SAT subject test benchmarks prior to retake
 - Encourage districts to utilize current ACT or SAT tools, training, and support
 - Partner with the Tennessee State Board of Education to determine appropriate interventions for a student who does not meet necessary benchmarks for postsecondary success
12. The department should ensure annual tests provide **clear reports** for educators, students, and parents that point to alignment to postsecondary readiness.
 - Already created sample reports for feedback
 - Already shared samples with assessment task force for feedback
 - Share with additional groups, including parent advisory group noted in recommendation nine

Test Scheduling and Logistics

13. The department should convene a testing scheduling and logistics advisory group that is representative of district personnel across the state to **address logistical challenges**. This team should focus on test scheduling that minimizes disruptions and provide ongoing guidance and

support (in the form of sample schedules and process guides) particularly during the first year of TNReady.

- Already convened Testing Scheduling and Logistics Advisory Group
- Already shared initial test schedule samples with assessment task force
- Share test schedule samples and guidance with districts
- Provide districts with ongoing technical assistance
- Analyze results for continued improvement

Other

14. The department should create **additional portfolio accountability options** for teachers in non-tested grades and subjects, specifically for first grade, from which districts can choose.
 - Already started first grade portfolio process
 - Begin pilot in spring 2016
 - Examine pilot results in preparation for statewide roll-out during 2016-17
15. The department must ensure future and new teachers have explicit knowledge about assessment for learning. Tennessee **teacher preparation programs** should include a specific curriculum or module on assessment with specifics to Tennessee. All new teachers, whether in tested subjects or not, also need specific training with embedded professional learning around this area of the TEAM rubric.
 - Create advisory group, including higher education faculty, district administrators, and educators, to determine contents of TN-specific assessment for learning module
 - Determine policy and process changes with the Tennessee State Board of Education
 - Pilot assessment for learning module with various teacher preparation programs during fall 2016
16. The department should work directly with districts to increase awareness of the realities of **test anxiety** while providing specific guidance in how to help educators avoid passing on stress or test anxiety to students. School counselors must be available to assist in this work during test administration.
 - Communicate expectations of assessment principles to directors at Superintendent's Study Council Conference in September
 - Provide administrators and school counselors with additional information and training on how to proactively create positive testing environments and how to work with students experiencing test anxiety

Conclusion

The task force believes that the principles and recommendations presented in this report are based on a comprehensive review of the state of assessment in Tennessee. After considering the primary areas of focus, including annual standardized assessments, district formative assessments, and test preparation and logistics, it is the view of the task force that if these principles are internalized and recommendations completed, assessment can yield what it is intended to do. The task force believes that assessment can and should become part of the normal, ongoing teaching and learning cycle. In doing so, assessment is viewed not as a competing interest but as part of the essential fabric needed to achieve high academic expectations for all students in Tennessee.

The task force recommends that the commissioner and department staff begin detailing specific owners for each recommendation and accompanying action steps, as well as the associated timeline to ensure completion. In addition, the task force desires to be continually engaged in reviewing new information as it is available about the state of assessment, especially since this is a year of transition with new statewide assessments in English language arts, mathematics, and social studies. The task force recommends that these principles and recommendations be revisited annually to improve the environment around assessment, maximize the benefits, and limit challenges to all stakeholders.

Appendices

Appendix 1: Agendas and PowerPoints from each meeting

Appendix 2: Minutes from each meeting

Appendix 3: Articles read for meetings

Appendix 4: Other state reports reviewed

Appendix 1

Agendas & PowerPoints



Assessment Practice Task Force: Meeting One

April 6, 2015

Agenda:


1. Introductions
2. Overview of Purpose - Commissioner McQueen
3. Discussion
 - a. Are there additional goals from this group?
 - b. What is your vision for an "end product"?
4. Overview of State Assessments – Emily Freitag and Nakia Towns, DOE
 - a. Assessment purposes, definitions and history
 - b. Current state assessments overview
5. Goals for Studying District Assessment Practices – Alyssa Van Camp, SCORE
 - a. Current district-level survey draft of what we plan to study
 - b. Teacher and student-level surveys
6. Discussion
 - a. Feedback on district-level survey draft
 - b. Ideas for teacher and student-level surveys
7. Closing & Next Steps

Documents:

1. Agenda with dates and HW questions on logo
2. Press Release
3. Task Force Contact Information
4. TDOE PowerPoint: Assessment overview
5. SCORE PowerPoint: District Survey
6. 2014-2015 Testing Schedule
7. HW: Reading: A framework for Considering Interim Assessments
8. HW: Stakeholder Interview Guide

Future Meeting Dates:

- Wednesday, May 20 (full day)
- Tuesday, June 16 (morning)
- Wednesday, July 15 (morning)



Assessment Practices Task Force

Overview of Purpose

0

Goals

- To conduct an environmental scan of assessment usage and practices across the state
- To establish principles addressing purposes and goals of state assessment relative to locally designed or locally chosen assessments (i.e. formative assessments)
- To define appropriate practices associated with these principles that best support decision-making at the state, district, school and teachers levels
- To gain insight on ways to best communicate about TN Ready to all stakeholder groups

EDUCATION 1

Discussion


- Are there additional goals from this group?
- What is your vision for an end product?

EDUCATION 2

Today's Goals

- Review the landscape of assessment in Tennessee
- Review district survey
- Gain feedback on district survey and and teacher/student surveys

EDUCATION 3



Assessment Practices Task Force

Overview of Assessment in Tennessee

Agenda

- Overview of assessment purposes, definitions and types
- Review of assessments:
 - Source
 - Purpose
 - Administration
 - Time
 - Cost

EDUCATION 5

Assessment Purposes

- To measure student mastery of a set of knowledge
- To identify what students know and do not know
- To compare performance
- To place students on a spectrum of performance
- To measure student growth
- To baseline current student understanding
- To measure performance relative to a goal
- To measure readiness
- To plan instruction
- To monitor effectiveness of interventions

Assessment Definitions

Key definitions:

- **Validity: how well a test measures what it is purported to measure**
- **Reliability:** Is the degree to which an assessment tool produces **stable and consistent** results
- Skill based assessment document student performance of a specific skill (i.e. reading comprehension). Standard based assessments measure student mastery of a specific standard.

Criterion versus Norm Referenced

Dimension	Criterion-Referenced Tests	Norm-Referenced Tests
Purpose	To determine whether each student has achieved specific skills or concepts	To rank each student with respect to the achievement of others in broad areas
Content	Measures specific skills which make up a designated curriculum	Measures broad skills sampled from a variety of texts, resources, and the judgments of curriculum experts
Item Characteristics	Each skill is tested by at least four items to test performance and minimize students guessing	Items are selected that discriminate between low and high achievers
Score Interpretation	Each individual is compared with a preset standard	Each individual is compared with other examinees and assigned score

Assessment Categories

- National
- State
- District
- Classroom

National Assessments

- National Assessments:
 - ACT / SAT
 - EXPLORE / PLAN
 - NAEP
- For each assessment, we will share:
 - Source
 - Purpose
 - Administration
 - Time
 - Cost

ACT/SAT

- Source: State Requirement in accordance with TCA 49-6-6001(b)
- Purpose:
 - To measure college readiness
 - Used to determine HOPE eligibility
- Administration
 - All 11th graders each year – ACT (No districts administer SAT in school.*)
- Time:
 - English: 45 minutes
 - Math: 60 minutes
 - Reading: 35 minutes
 - Science: 35 minutes
 - Writing: 30 minutes
- Cost: ~\$45.25/pupil**

* Students may choose to participate in SAT in a non-district administration using a voucher. Or students may use a voucher to participate in ACT in a non-district administration.
 ** Average cost between ACT and SAT, funded via BEP based on number of graduating students.

Explore / Plan

- Source: State Requirement in accordance with T.C.A. 49-6-6001(b)
- Purpose: To gauge college readiness.
- Administration:
 - Explore: all 8th graders annually
 - Plan: all 10th graders annually
- Time:
 - Explore: 120 minutes
 - Plan: 1 hour, 55 minutes
- Cost: ~\$10.50/pupil*

* Cost will increase to \$12.50/pupil in 2015-16

NAEP

- Source: Federal requirement in accordance with No Child Left Behind Act 1111(c)(2) & National Assessment of Educational Progress Authorization Act (Public Law 107-279 III, Section 303)
- Purpose:
 - Common metric across states to measure relative performance
 - Continuous assessment to allow for progress monitoring over time
- Administration:
 - Sampled administration - In an average state, 2,500 students in approximately 100 public schools are assessed in each grades 4 and 8, for reading and mathematics. Typically, 30 students per grade per subject are selected randomly in each school.
 - Administered every two years, math and ELA in 4th and 8th
 - Periodic administration of subject specific tests in US History, writing, the arts, civics, economics and technology
- Time: 60 minutes
- Cost: \$0

State Assessments

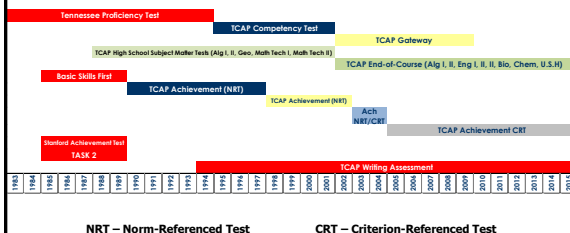
- State Assessments:
 - TCAP (this year)
 - Achievement
 - End of Course
 - Writing
 - TCAP (next year)
 - TNReady (ELA and Math)
 - Achievement/EOC (Science and Social Studies)
 - TCAP Alternate Assessments
 - Portfolio for 1%
 - ACCESS for ELLs

History of State Assessments

- Tennessee Comprehensive Assessment Program** is the state's summative assessment system
 - Commenced with Tennessee Proficiency Test in **1983**
- Mandated by state legislature with the **1992** Education Improvement Act (EIA)
 - Tennessee Value-Added Assessment System (TVAAS)** launched via 1992 EIA
- Various components of test were made and introduced from **1999 until 2002**
- CTB/McGraw-Hill and Measurement, Inc.** held main testing contract in early years; now, **Pearson and Measurement, Inc.**

History of Assessment in Tennessee

Tennessee Comprehensive Assessment Program (TCAP) Timeline: 1983-2015



NRT – Norm-Referenced Test

CRT – Criterion-Referenced Test

TCAP (Achievement, End of Course and Writing)

- Source: Federal law NCLB Act of 2001 1111(b) and T.C.A. 49-1-602
- Purpose:
 - Measure mastery of grade level (or course) standards
- Administration
 - Achievement: Grades 3-8, Math, ELA, Social Studies and Science
 - EOC: Algebra I/II, English I/II/III, Biology, Chemistry, US History
 - Writing: Grades 3-11
- Time: (See following slides.)
- Cost:
 - Achievement: ~\$24.36/pupil (includes ELSA)
 - EOC: ~\$20.33/pupil
 - Writing: ~\$13.31/pupil

Achievement Testing Times 2014

Content Area	Average Testing Time in Minutes
TCAP Achievement - Reading Language Arts	124
TCAP Achievement – Mathematics	78
TCAP Achievement – Science	70
TCAP Achievement – Social Studies*	98

End of Course Testing Times (untimed)

Content Area	Average Estimated Testing Time in Minutes
English I, II, and III	120-240
Algebra I and 2	120-240
Biology	120-240
Chemistry	120-240
U.S. History	120-240

Writing

Content Area	Average Testing Time in Minutes
Grades 3-5	90
Grades 6-8	120
Grades 9-11	120

TCAP Changes in 2015-2016

- Source: Federal law NCLB Act of 2001 1111(b) and T.C.A. 49-1-602
- Changes:
 - Math, RLA and Writing will be replaced by TNReady
 - Adds Geometry and Integrated Math I, II, and III for secondary
 - Science Achievement and EOC assessments will continue
 - Social Studies Achievement and EOC assessments will be new
- Time: (estimates)
 - Math: 3 subtests (1 in February, 2 in April) 60-90 minutes each
 - ELA: 4 subtests (2 in February, 2 in April) 60-90 minutes each
- Cost:
 - TNReady: ~\$21.14/pupil*

* Cost assumes same number of test takers from spring 2014 administration.

TCAP-Ait (Portfolio will be replaced by NCSC)

- Source: Authorized in regulation 34 C.FR Part 200. IDEA, section 504 of the Rehabilitation Act of 1973 and title I require inclusion of all students with disabilities in the state assessment system.
- Purpose:
 - To measure student growth over the course of a school year.
- Administration:
 - Teachers collect data across the year through the end of January.
 - Administered to up to 1% of students excluded from the general state assessment.
- Cost:
 - ~\$127.61/pupil*

* Cost for 2015-16 NCSC pending RFP. NCSC will only include RLA and Math. TCAP-Ait will continue for Social Studies and Science.

WIDA/ACCESS for English Learners

- Source:
 - Administered in accordance with NCLB 2001 1111(b)(3)(c)(ix)(III)
- Purpose:
 - To evaluate English proficiency
- Administration:
 - Administered to students in their first year enrolled in a U.S. school until they demonstrate English language proficiency.
- Time:
 - Listening: 20-25 minutes
 - Reading: 35-40 minutes
 - Writing: Up to 1 hour
 - Speaking: Up to 15 minutes
- Cost:
 - ~\$25.75/pupil*

* Cost assumes same number of ELDA test takers from spring 2014 administration.

District Assessments

- K-2
- RTI2 Screeners
- Other District Assessments

K-2

- Optional assessment: SAT 10
- District participation trends
 - 2013 - 117 districts (168,679 students)
 - 2014 - 88 districts (124,512 students)
 - 2015 - 76 districts (77,730 students based on current orders)
- Purpose:
 - To measure student mastery in early grades
 - To inform intervention and remediation around numeracy and literacy
 - Can be used for value added for K-3 teachers (after baseline year)
- Norm-referenced
- SAT10 will be phased out after 2015-16
- Potential RFP for 2nd grade option
- Cost: ~\$20.10/pupil

RTI² Universal Screeners

- Response to Instruction and Intervention is a problem-solving approach to identifying and addressing student skill deficits.
- Universal screeners are not assessments in the traditional sense. They are meant to be quick, informative, and non-intrusive and are given 3 times a year to identify at risk students.
- The most common screeners selected and used by districts in grades K-5 are:
 - AimsWeb (30)
 - EasyCBM (15)
 - STAR 360 (47)
- In most of these programs, reading fluency is assessed for one minute with individual students while math computation is group administered and takes 5-8 minutes. This amounts to less than 30 minutes per child throughout the year.

Other District Assessments

- Terms Used:
 - Benchmark
 - District Interim
 - Common formative
 - Local EOCs
 - Honors exams
- Types:
 - Purchased
 - District-created
 - Teacher-written

Questions



Future Meeting Dates

- May 20 – full day
- June 16 – morning
- July 15 – morning

Next Steps

- Homework
- Goals for May 20th meeting
 - Review results of surveys
 - Review usages of assessment results
 - Design principles for purposes and goals of assessment
 - Begin drafting recommendations



District Assessment Survey

- What assessments are districts in Tennessee currently implementing?
- How are these assessments used in districts in Tennessee?
- How much time is spent taking assessments in Tennessee?





Tentative Timeline

- April: District, principal, teacher, and student surveys administered.
- May: District interviews and principal, teacher, and student focus groups conducted.
- June: Data analysis and report writing.
- July: Report review and release.

A collage of two images. The top image shows a student wearing a white head-mounted display (HMD) and looking through a microscope. The bottom image shows a woman with long brown hair looking at a tablet computer.

Assessment Practices Task Force: Meeting Two

Task Force Goals:

- To conduct an environmental scan of assessment usage and practices across the state.
- To establish principles addressing purposes and goals of state assessment relative to locally designed or locally chosen assessments (i.e. formative assessments).
- To define appropriate practices associated with these principles that best support decision-making at the state, district, school and teachers levels.
- To gain insight on ways to best communicate about TN Ready to all stakeholder groups.

Meeting Two Goals: (We will...)

- Discuss findings from surveys and interviews
- Review state assessments and current performance and use
- Conduct a deep dive of the plan for TNReady
- Draft initial principles for state and local assessments

Agenda:

Topic	Components	Timing
Welcome	<ul style="list-style-type: none"> • Introductions • Task Force Goals • Review the Agenda, Goals and Norms 	9:00-9:15
Discussion of Homework	<ul style="list-style-type: none"> • Discussion of Article • Discussion of findings from interviews 	9:15-10:00
Survey Results	<ul style="list-style-type: none"> • Preliminary Findings • Discussion of findings and places for additional focus 	10:00-11:30
Pick Up Lunch		11:30-11:45
State Assessments: Goals, Current Results and Uses	<ul style="list-style-type: none"> • Overview of Goals in Strategic Plan • Current Results • Case Studies of Research 	11:45-12:30
Overview of TNReady	<ul style="list-style-type: none"> • Review of Sample Items • Content Design • Logistics Plans • Timeline for results • Reports planning process • Communications discussion 	12:30-1:45
Break	<ul style="list-style-type: none"> • Move three facilitated groups during break 	1:45-2:00
Small groups facilitated discussion of principles for state assessments	<ul style="list-style-type: none"> • Discussion of principles – draft purposes and goals of state assessment relative to local assessments 	2:00-3:15
Recap & Overview Homework	<ul style="list-style-type: none"> • Next steps • Homework 	3:15-3:30

Future Meeting Dates:

- Morning of Tuesday, June 16
- Morning of Wednesday, July 15



Strategic Plan Draft

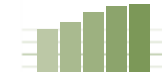
April 2015

Tennessee has made major strides in improving educational outcomes.

Elementary and Middle Schools



Fastest improving state in the nation on 4th and 8th grade NAEP



Consistent gains on TCAP every year since new assessments in 2010

High Schools



Fastest growing graduation rate of any state



ACT statewide average has increased to 19.3

At the same time, the world has changed and today's students need much more to be able to succeed.

By 2025, 55% of all new jobs will require postsecondary education



Postsecondary graduates are more likely to be employed and have higher earnings than high school graduates.

The gaps in employment and earnings between these groups have grown substantially over time.

Tennessee students are struggling in the early years after high school.

72,865 Students

2007 Cohort of High School Freshmen



10,545 students did not graduate from high school



22,334 students graduated from high school and entered the workforce and earn an average salary of \$9,030 annually



40,235 students enrolled in postsecondary.

58 percent were still enrolled in one year (or 20,418 of the 35,055 who enrolled immediately after graduation).

3,514 had completed a certificate or degree within three years.

Tennessee Promise gives students an incredible, new opportunity.

Free, Public K-14 System

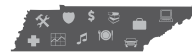
Grades K-12

Grades 13-14

Tennessee Promise

Additional Postsecondary Education and Career Opportunities

It's now our responsibility to set students up for success.



Changing World



Progress

Given our progress, the changing world, and the opportunity of Tennessee Promise, we must reorganize around a new vision:




Grades K-12

Grades 13-14

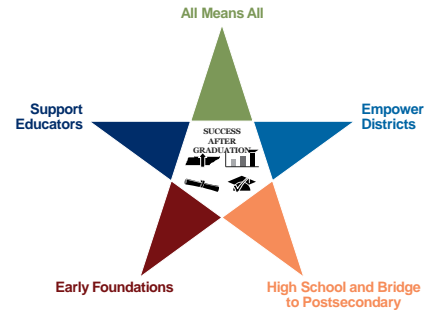
TN Promise

SUCCESS AFTER GRADUATION

To ensure our students are ready for postsecondary success, we must meet the following goals.

SUCCESS AFTER GRADUATION		
<p>GOAL #1</p>  <p>Tennessee will continue its rapid improvement and rank in the top half of states by 2019.</p> <p>MEASUREMENT We will rank in the top half of states on 4th and 8th grade NAEP in 2019.</p>	<p>GOAL #2</p>  <p>The average ACT score in Tennessee will be a 21, allowing more students to earn HOPE scholarships.</p> <p>MEASUREMENT Tennessee will have an average public ACT composite score of 21 by 2020.</p>	<p>GOAL #3</p>  <p>A majority of high school graduates will go on to earn a certificate, diploma, or degree.</p> <p>MEASUREMENT The class of 2020 will be on track to achieve 50% postsecondary completion within six years.</p>

To achieve success after graduation, we will organize our work around the following strategic priorities.



Early Foundations

Ensure students are building the necessary skills in early grades to be ready for future success

- ★ Tennessee currently spends more than \$250 million annually on pre-kindergarten, Head Start, and state-subsidized child care, but we do not know whether these students are ready for kindergarten upon completion. We will work to strengthen and measure Pre-K pathways to increase the number of students ready for Kindergarten.
- ★ Kindergarten through second grade are crucial, yet we lack a strong measure of student progress and adult impact.
- ★ Third grade reading levels are highly predictive of both K-12 and postsecondary success. We must increase the number of students who end third grade on grade level and have a common definition for what that means.
- ★ Teachers are not receiving the preparation they need to be successful teaching foundational literacy skills in grades K-3. We will work with teacher preparation programs to strengthen and align their support for pre-service teachers in this area. Priority will be placed on improving pre-service teacher reading standards.

Support Educators

Strategically support the preparation and development of a strong educator workforce in Tennessee

- ★ All teachers have performance evaluations, but many continue to lack access to effective tools to get better. We must ensure teachers have access to high-quality, job-embedded professional development that is aligned to their performance evaluation.
- ★ Teachers and leaders need to come out of preparation programs with the tools needed to equip students to meet higher standards, and prepare them with the skills they will need to compete in today's workforce. We must work to better align K-12 and higher education and strengthen their partnerships with local districts.
- ★ Teachers lack opportunities to advance their careers. We will work to strengthen pipelines for teacher leaders.
- ★ Student attendance and other non-academic factors have a significant impact on academic achievement. We can better leverage and coordinate our efforts to help address these factors and support educators in advancing student learning.

Empower Districts

Provide districts with the data, support, and autonomy they need to make the best decisions for their students

- ★ Tennessee's districts and schools vary greatly in size, demographics, opportunities and challenges. We must prioritize flexibility and choice to ensure that we meet the needs of all districts.
- ★ The lens through which districts exercise choice and prioritize options is through a thorough strategic planning process. We will work to ensure that districts complete one district-wide comprehensive plan.
- ★ Districts are best positioned to make informed decisions about how to manage and support their schools and educators. We should highlight their effective practices by creating a culture of sharing and collaboration, and expanding opportunities for networking.
- ★ High performing districts should have more opportunities for earned autonomy.
- ★ Tennessee has a wealth of data at its disposal. We should better support districts in utilizing this data by providing resources to guide decision making around specific data sets.

All Means All

Provide individualized support and additional opportunities for students who are furthest behind

- ★ A majority of Tennessee students are economically disadvantaged, and large numbers are members of other high need groups. Tennessee cannot succeed as a state unless these students are successful. We will support districts in networking around what works and sharing best practices to meet the needs of all students.
- ★ Economically disadvantaged and African American students made massive strides in the last several years. However, major gaps still exist for these students and for students with disabilities and English learners. We must encourage equity by providing data transparency around these gaps.
- ★ Large numbers of students remain stuck in failing schools. While the ASD and iZones show promise in changing these schools' trajectories, more—and faster—intervention is needed. We will encourage innovation and collaboration among schools in the bottom 5% of the state.
- ★ Students who are farther behind must grow at a faster rate than their peers to catch up. However, we know they will need more support to do so. We must work to ensure an equitable allocation of resources and coordinating additional wrap-around supports for them where possible.



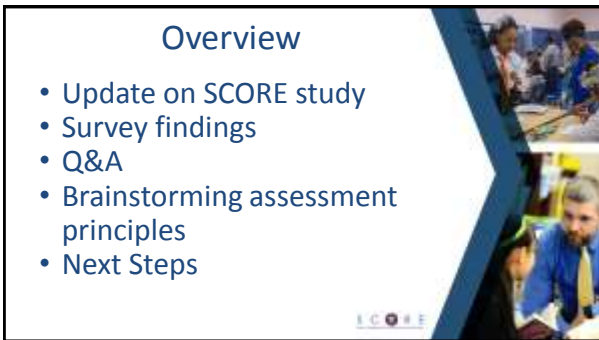
High School and Bridge to Postsecondary

Ensure high schools prepare significantly
more students for postsecondary

- High schools have become much stronger at graduating students, but lack the data to track students beyond graduation.
- High schools are not advancing student skill levels sufficiently; eighth grade scores almost fully predict 12th grade outcomes. We will work to provide better data to schools to support the transition from eighth to 12th grade and to ensure students continue to progress in high school.
- Many students have insufficient access to rigorous courses that prepare them for postsecondary and are aligned to workforce needs.
- The handoff between high school and postsecondary is weak. Too many students lack the necessary support and guidance to negotiate this transition. We must provide better guidance to districts and schools on how to leverage the role of school counselors in supporting this transition.
- To ensure that students are adequately prepared for postsecondary, we will work to ensure alignment of our current standards to ACT benchmarks.




SCORE
Assessment
Study
 May 20, 2015



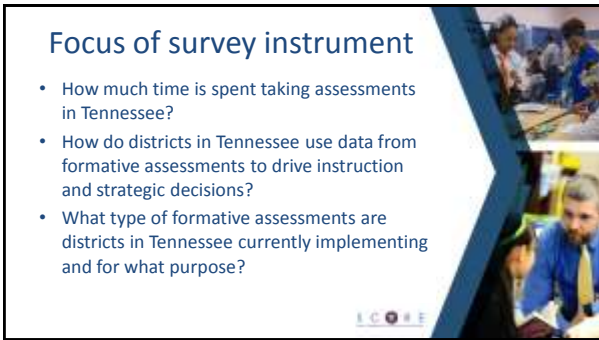
Overview

- Update on SCORE study
- Survey findings
- Q&A
- Brainstorming assessment principles
- Next Steps




Updates

- Administered teacher, principal, and district surveys (March-April 2015)
- Analyzed survey data (April 2015)
- Launched focus groups with teachers and principals (April-June 2015)
- Preparing for district interviews (May 2015)



Focus of survey instrument

- How much time is spent taking assessments in Tennessee?
- How do districts in Tennessee use data from formative assessments to drive instruction and strategic decisions?
- What type of formative assessments are districts in Tennessee currently implementing and for what purpose?



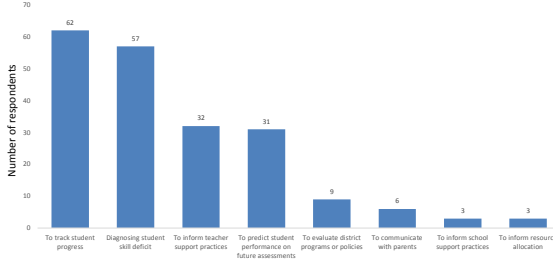
Response Rates

Teacher Survey	Principal Survey	District Survey
10,000 responses (approx.)	200 responses (approx.)	69 Responses
14.9% response rate	11.1% response rate	48.9% response rate

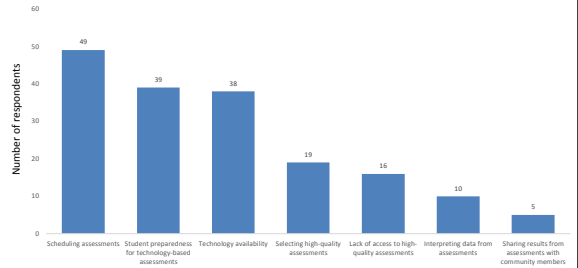


District Survey Results

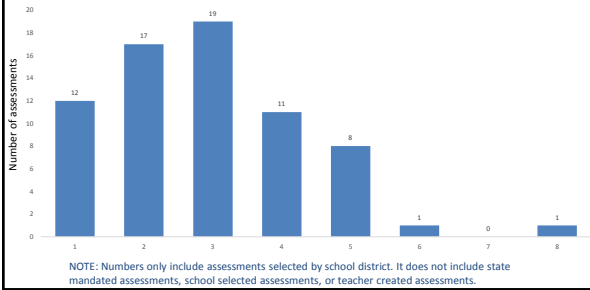
What are the top three ways you use assessments in your district?



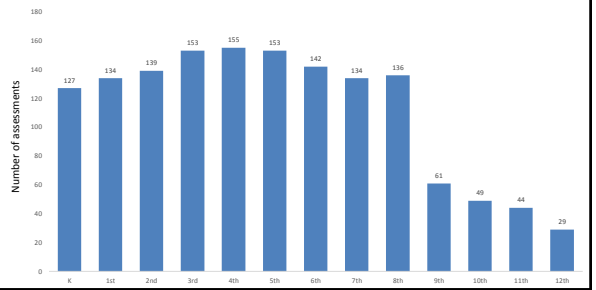
What are the top three challenges you face with assessments as a district?



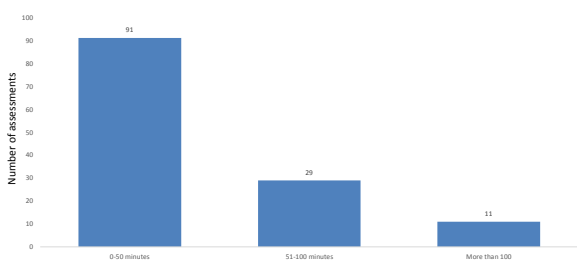
Number of district-selected assessments administered by school district



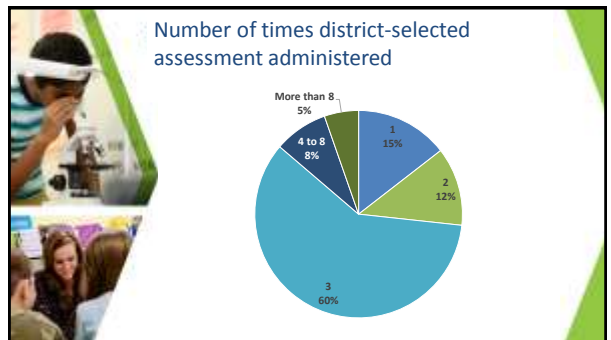
Grades Tested



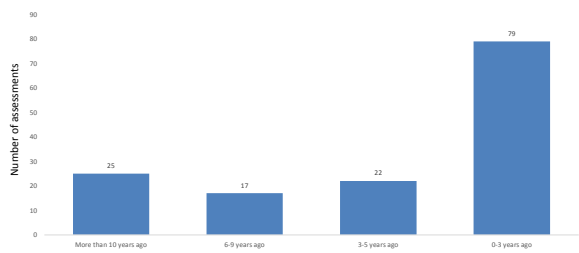
Amount of time students spent taking district-selected assessments



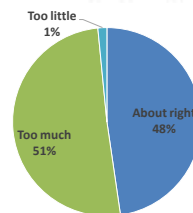
Number of times district-selected assessment administered



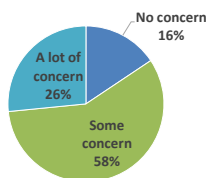
What year did your school begin administering this assessment?



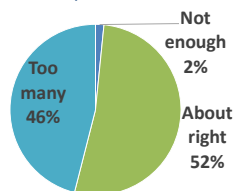
Do you think the time spent on assessment in this district is too much, too little, or about right?



What level of concern do you have about the time spent on assessments in your district?



Do you think students in this district are given **too many** assessments, **not enough** assessments, or **about the right number** of assessments throughout the school year?

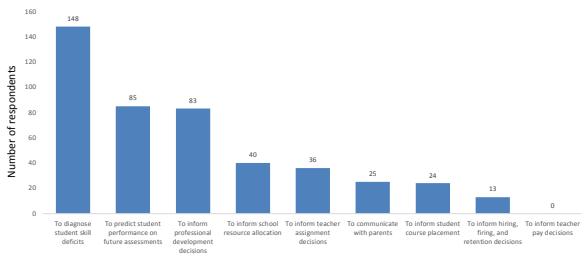


"It is one of those things that you can't live with and can't live without...When we started out with formative assessments in our district, we quickly learned that one size does not fit all. We had to create some of our own. You have to have the assessments though because if you are going to have the data conferences monthly and use it for grouping and intervention, you need that information. It is a real balancing act though."

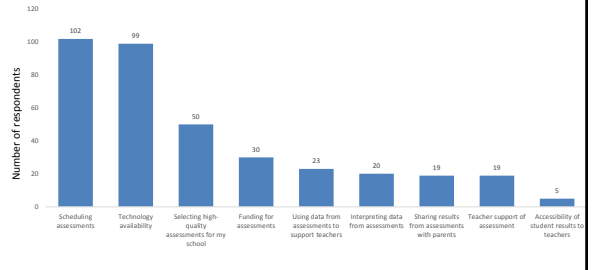
-Superintendent

Principal Survey Results

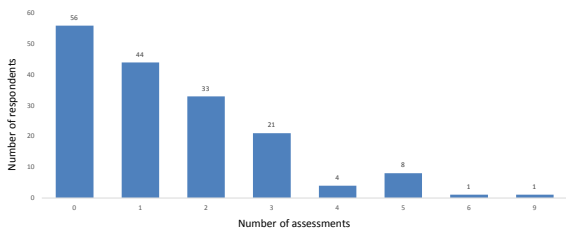
What are the top three ways you use assessment as a principal?



What are the top three challenges you face with assessment as a principal?

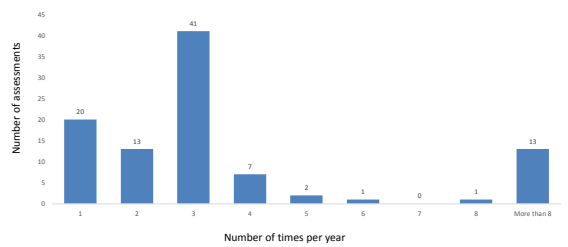


How many assessments do you implement that are **unique to your school**?

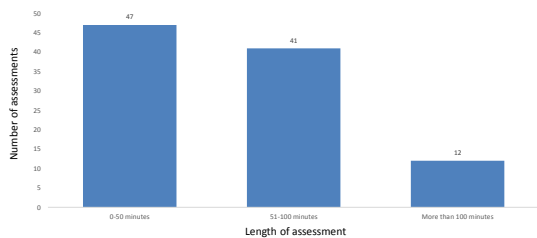


NOTE: Numbers only include assessments selected by school district. It does not include state mandated assessments, school selected assessments, or teacher created assessments.

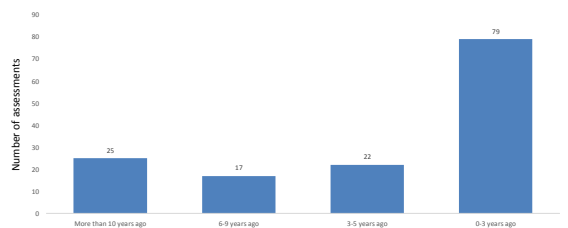
How many times per year is this assessment given?



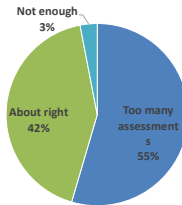
How long do students spend taking this assessment?



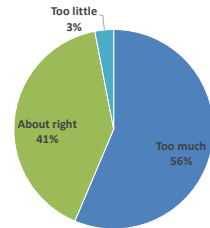
What year did your school begin administering this assessment?



Do you think students in this district are given **too many** assessments, **not enough** assessments, or **about the right number** of assessments throughout the school year?



Do you think the time spent on assessment in this school is too little, about right, or too much?

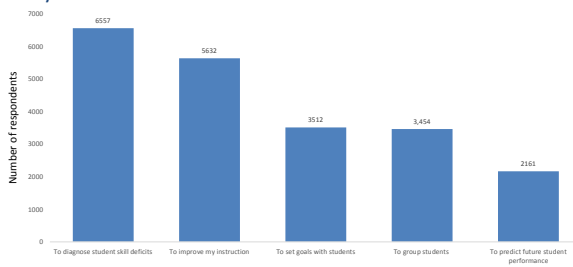


"I think formative assessments are essential. Teachers are required to use formative assessments but they also want to use them. Teachers can see instant results and it drives their instruction. When teachers were able to personalize student learning, because they had an intimate relationship with that data, it helped not only that child, but the entire class."

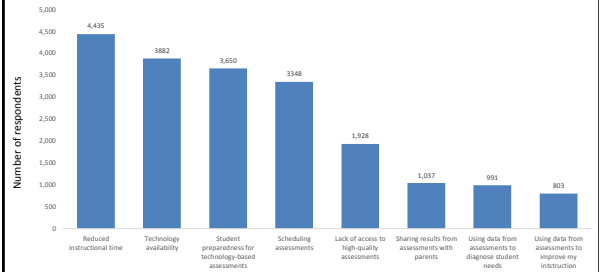
—Assistant Principal

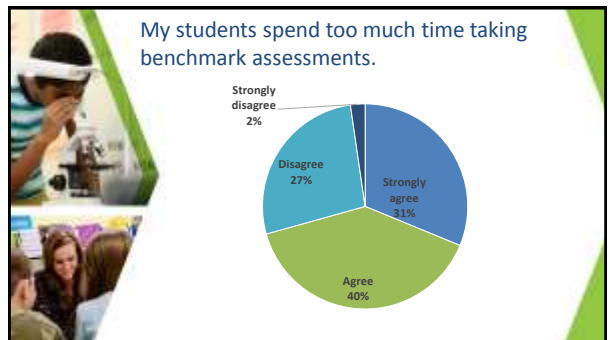
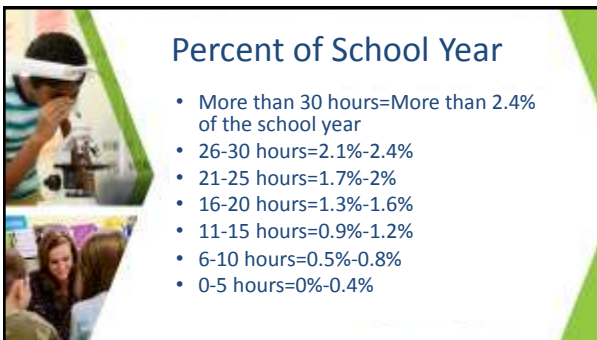
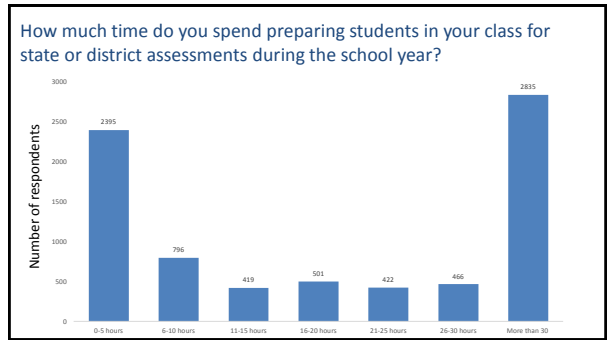
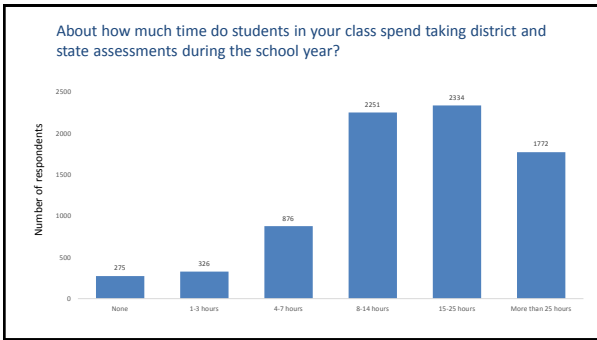
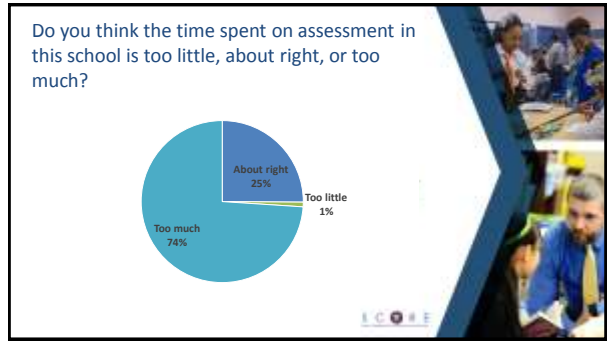
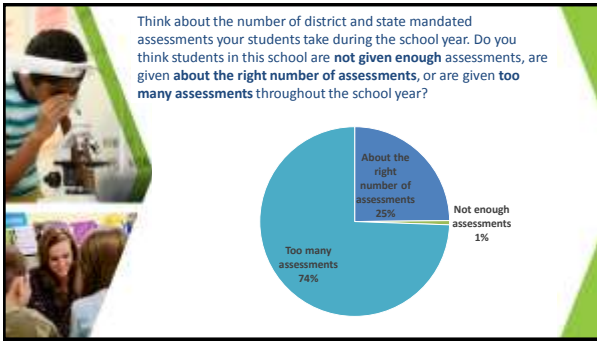
Teacher Survey Results

What are the top three ways you use assessment in your classroom?

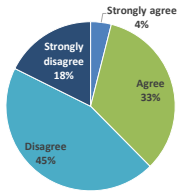


What are the top three challenges you face with assessments as a teacher?

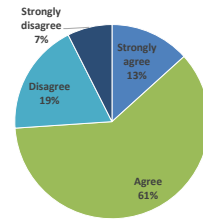




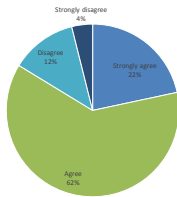
Overall, the benefits to my students from benchmark assessments are worth the investment of time and effort.



Benchmark assessment results are available to me in a timely manner.



I understand how to use results from benchmark assessments to improve my teaching.



Takeaways

- Similarities between principal and district responses
- Teacher results indicate they believe they spend too much time on testing
- Consistent uses of assessment across stakeholders
- Logistics of implementation are the biggest challenges
- Mixed response on benchmark assessments

Q&A



Small Group Discussions

- Based on the survey findings and your own experience, what are the top challenges this stakeholder group faces with assessment? What are the top uses of assessment for this group?
- Based on the survey findings and your own experience, what are possible solutions to these challenges?
- Based on the survey findings and on your own experience, brainstorm a few principles and norms for assessment practices in Tennessee.

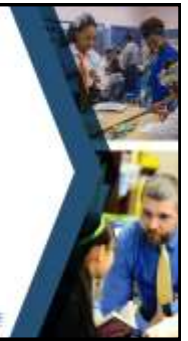
Small Group Discussions

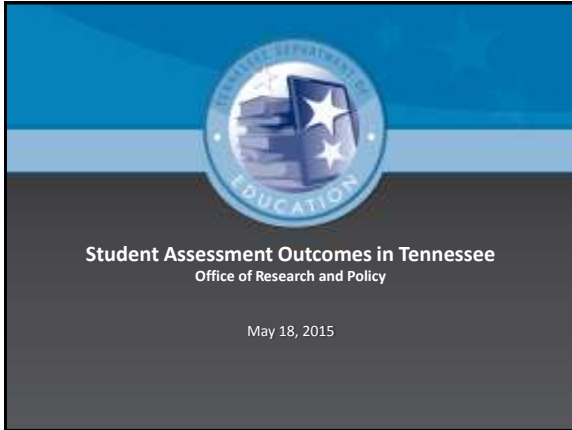
Teachers	Principals	Districts
Val Love	Sharon McNary	Susan Lodal
Jasmine Carlisle	Beth Unfried	Nancy Ashe
Becky McBride	Sara Heyburn (Nathan)	Mary Reel
Phillip Eller	Emily Freitag	Wanda Shelton
Vicki Kirk	Kathleen Airhart	Mike Winstead
Virginia Babb	Chairman Brooks	Nakia Towns
Chairman Gresham	Drew Jacobs	Chairman Forgety
Commissioner McQueen	Stephen Smith	Alyssa Van Camp



Next steps

- May-June: Focus groups with teachers and principals
- May-June: District interviews
- June: Data analysis
- July: Report release





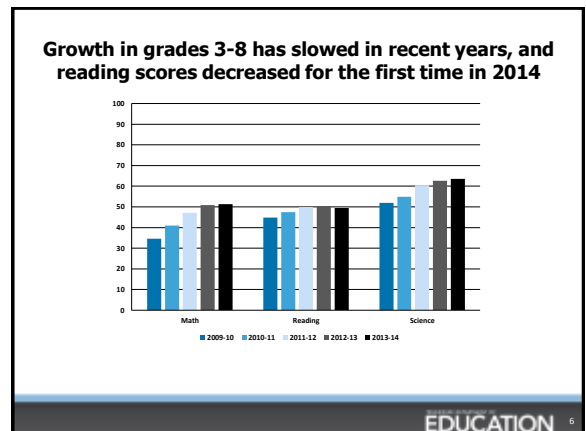
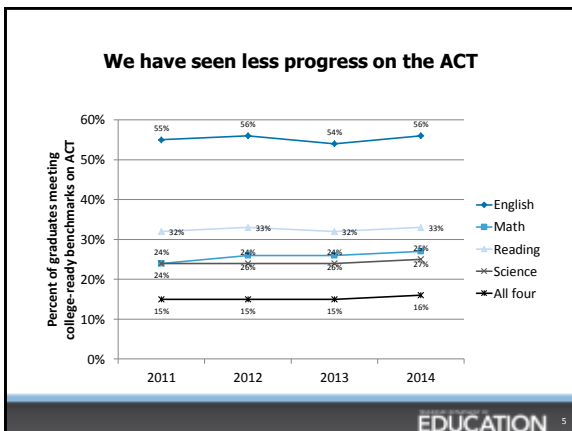
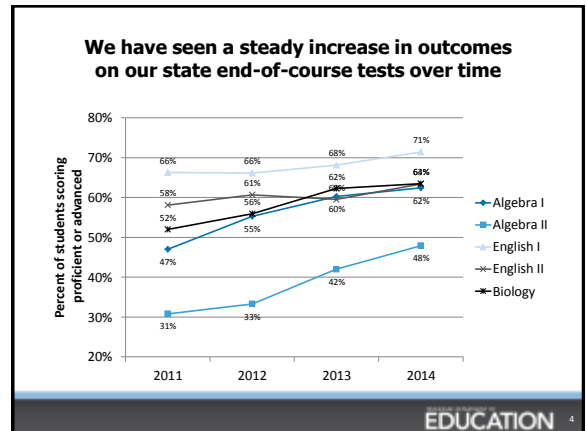
Landscape of Student Testing in Tennessee

	TN Only	TN & Other States
K-2		SAT-10 (some districts)
3 rd - 8 th	TCAP Achievement <ul style="list-style-type: none"> Reading/Language Arts+* Math+* Science* Social Studies Writing Assessment	NAEP (4 th & 8 th) <ul style="list-style-type: none"> Reading/Language Arts Math EXPLORE (8 th) <ul style="list-style-type: none"> English Math Reading Science
9 th - 12 th	TCAP End of Course <ul style="list-style-type: none"> English I, II, & III+* Algebra I & II+* Biology* & Chemistry U.S. History Writing Assessment - except grade 12	PLAN (10 th) - same as EXPLORE ACT (11 th) - same as EXPLORE NAEP (12 th) <ul style="list-style-type: none"> Reading/Language Arts Math

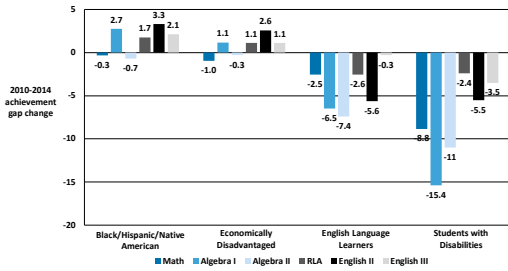
+Used for district accountability *Used for school accountability

We need to prepare students to succeed in post-secondary work

- College Readiness**
 - In 2014, 16% of TN high school seniors hit the ACT college readiness benchmark in all four subjects compared with 26% nationally.
- Remediation**
 - Nearly 70% of TN college freshmen require remedial coursework during their first year of college.
- Educational Attainment**
 - Tennessee ranks 43rd in working adults with a 2-year degree or higher.



Between 2010 and 2014, ELL and SWD gaps have grown considerably although ED and minority gaps have narrowed



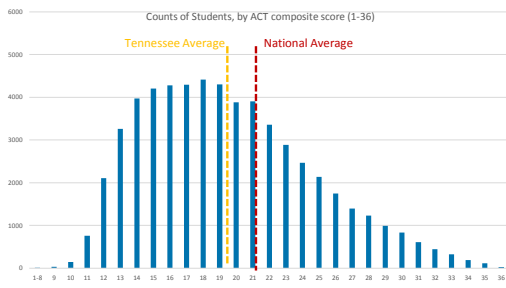
Change in gap size is measured between 2010 and 2014 for all subjects except A2 and E3. For these subjects, change is measured between 2012 and 2014.

Digging in further: A few examples

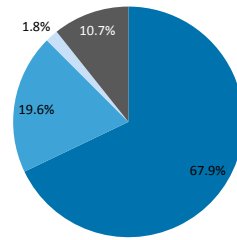
- Example 1: ACT performance
- Example 2: Writing practices
- Example 3: Chronic absenteeism



Tennessee's 2010 9th grade cohort's composite score falls about 1.6 points below the national average (20.9 to 19.3)

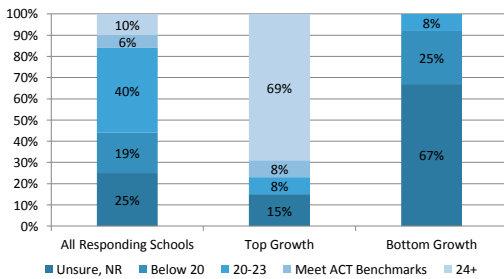


The majority of school principals say their school has an ACT goal and that it is a high priority

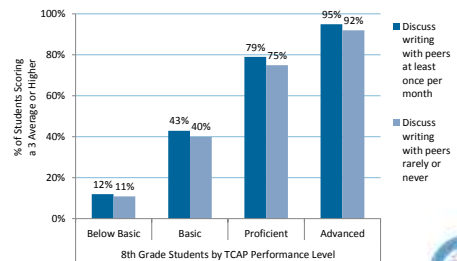


- School has goal that is high priority.
- School has goal that is not discussed.
- Not sure whether there is a school goal.
- No, there is no set goal.

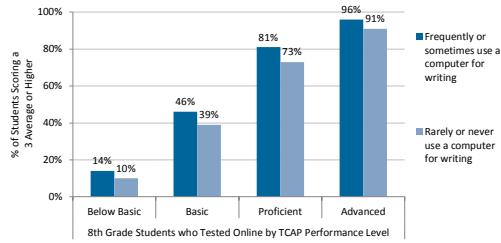
However, in bottom growth schools counselors are much less likely to know about a school goal, suggesting a failure to communicate about the goal and create a shared vision for ACT success



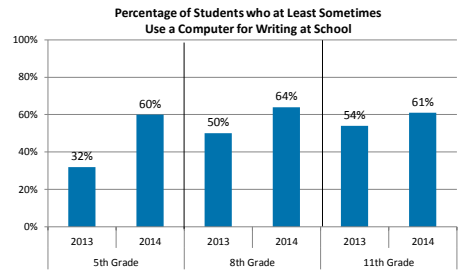
Writing Assessment scores vary meaningfully depending on classroom teaching practices – even among students who perform at the same level on TCAP



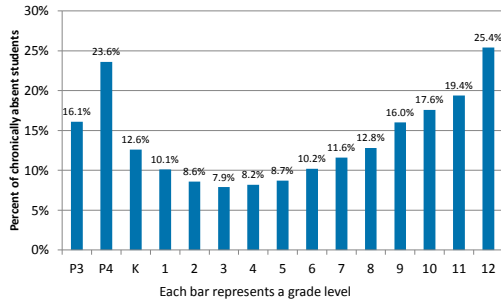
Whether or not students receive online writing practice also matters



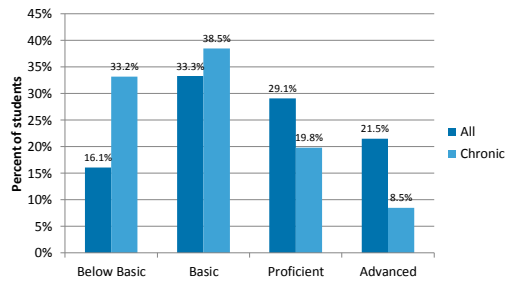
The percentage of students who used a computer for writing at school increased in 2014

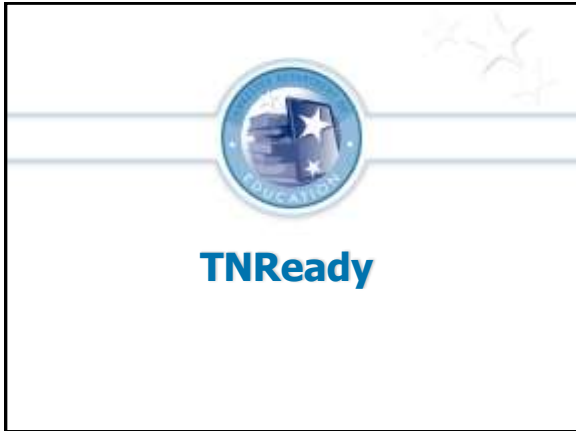


Tennessee's high attendance rate hides the fact that some students are missing a lot of school



Students who are chronically absent score proficient or advanced on math TCAP at lower rates than their peers





TNReady Overview

1. Goals of TNReady
2. Content Design & Sample Items
3. Logistics Plans & Technology
4. Reports & Timeline
5. Communications discussion

EDUCATION 2

Goals

3

Tennessee students are struggling in the early years after high school.

72,865 Students
2007 Cohort of High School Freshmen

 10,545 students did not graduate from high school	 22,334 students graduated from high school and entered the workforce and earn an average salary of \$9,030 annually	 40,235 students enrolled in postsecondary. 58 percent were still enrolled in one year (or 20,418 of the 35,055 who enrolled immediately after graduation). 3,514 had completed a certificate or degree within three years.
---	--	--

EDUCATION

Tennessee Promise gives students an incredible, new opportunity.

Free, Public K-14 System

Grades
K-12

➔

Grades
13-14
Tennessee Promise

Additional Postsecondary Education and Career Opportunities

EDUCATION

It's now our responsibility to set students up for success...

TNReady!

Changing World

Progress

Given our progress, the changing world, and the opportunity of Tennessee Promise, we must reorganize around a new vision:

TN Promise

SUCCESS AFTER GRADUATION

EDUCATION

TNReady: Goals

- Provide real information about student readiness for postsecondary work.
- Anchor expectations about what students need to know and be able to do at every grade level to be ready.
- Provide actionable information for parents and educators to support student growth.
- Provide information about our competitiveness as a state.

Math: The Big Picture

- We hear regularly from employers that our workforce lacks math skills. (Often basic skills.)
- The fastest growing sectors of our economy demand strong technical skills.
- The breadth of our assessments and the structure of past tests have led to instruction that races through skills, especially in early grades, without developing strong command of the basic skills and foundational concepts.
- *"Math people" are made. Not born.*

TNReady: Math Priorities

- 3-8: Focus on fewer concepts – assess those topics in a range of ways.
- HS: Strengthen coherence – assess topics in connected ways.
- Include authentic assessment of real-life situations.
- Support alignment with ACT.
- Include calculator-permitted and calculator-prohibited sections at every grade level.

ELA: The Big Picture

- We hear from postsecondary institutions and college students that students are unprepared for the demands of college, especially in reading analysis and writing.
- The gap between the text complexity of HS reading and college level reading, as of 2010, was a four year gap. Challenge meeting the demands of reading and writing are among the leading academic causes of college drop out.
- The segregation of reading and writing in previous assessments (across states) led to instructional isolation and a detrimental focus on discrete ELA skills to the neglect of preparation for the type of communications required in real world.

TNReady: ELA Priorities

- Integrate reading and writing skills in one score and in item design
- Focus on authentic workplace and postsecondary skills – reading, writing, editing
- Staircase complexity to prepare students appropriately
- Include a range of text types in reading and writing

Item Types Overview

Math Questions:

1. Equation
2. Graphic
3. Multiple Choice
4. Multiple Select
5. Performance Tasks (for grade 3-8 only)
6. Technology Enhanced Items

ELA Questions:

1. Writing tasks
2. Technology-enhanced (TEI)
3. Multiple choice
4. Multiple select
5. Evidence-based selected response

Calculator Policy

- Two central beliefs:
 - Calculators are important tools for college and career readiness
 - Students must be able to demonstrate many skills without reliance on calculators
- At all grade levels and in all courses, **TNReady will include both calculator permitted and calculator prohibited sections**
- Examples of permitted and non-permitted calculators, consistent with ACT and other benchmark assessments.
- Handhelds are permitted with online testing.
- Example for 5th Grade Standards – all will be available by the end of March

Practice Tools

TNReady Item
Sampler
(MICA)

TNReady
Practice Test
(MIST)

TNReady Item Sampler (MICA)

Purpose:

- To give educators access to questions like the questions on TNReady
- To give students the chance to practice with the tools and system in every day instruction

Timeline

- May – Teacher Access, 8-12 Items per grade and subject
- August – Student Access, 25-40 more items per grade and subject

Functions

What will teachers be able to do with MICA?

- Preview items with scoring information
- Build test-lets with available questions
- Assign tests to classes or individual students
- See reports on student performance on automatically scored items
- Score extended response student work
- Access practice tools

TNReady Practice Test (MIST)

Purpose

- To simulate a short-form of the TNReady test on the same platform as the operational test.

Timeline

- One practice test will be available for Part I and Part II throughout the year.
- It will be available on the testing platform (but it will not be available during operational testing windows.)
- The MIST Practice Test will be available during October, January and March.

Functions

What will schools be able to do with the practice test?

- It has all the same features and functions as the TNReady ready test, only the questions will be different.
- The set up process for the practice test is the same as the set up process for the operational test.
- Reports will be provided to the school on student performance on automatically scored questions.
- Teachers can score Extended Response questions using the same rubrics and scoring guides.

DEMO PRACTICE ITEMS



19

Logistics & Technology



20

Technology Specifications (pages 125-126)

- Device readiness is determined as follows:
 - K-5, 6-8, 9-12 Schools
1 device for every 2 students in the largest grade
 - K-8 Schools
1 device for every single student in the largest grade
- Traffic lights:
 - **Green** = 100% compliant
 - **Yellow** = 80 – 99% compliant
 - **Red** = less than 80% compliant

Technology Specifications (pages 125-126)

- Network readiness is based upon commonly accepted guidelines set forth by SETDA.
- Traffic lights:
 - **Green** = 50 kbps / testing student
 - **Yellow** = 20 kbps / testing student
 - **Red** = less than 20 kbps / testing student
- Remember that a read-aloud accommodations will require *150 kbps per student testing* with an accommodation.

Technology Specifications (pages 125-126)

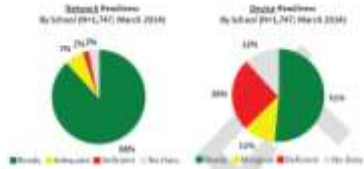
- General hardware requirements (OS independent)
 - 1 GB RAM (2 GB min recommendation)
 - 1024 x 768 screen resolution
(tablets must have a min 9.5 inch screen)
 - External keyboard and mouse
- OS requirements
 - Windows 7 or greater (Windows XP is dead!)
 - Mac OS X 10.6 or greater
 - iOS 6.x or greater
 - ChromeOS (2013 or later builds)

Technology Specifications (pages 125-126)

- Browsers
 - IE 9.x or later
 - Firefox 31 or later
 - Safari (version corresponding to OS X version)
 - Chrome 32 or later

Online Assessment Guidelines (pages 127-132)

- We believe that access to instructional technology is *fundamentally an issue of equity*. The guidelines are designed to encourage prioritizing these critical investments.
- Our goal is to have as *many* schools and districts participating in online assessment as is technically feasible based on local resources.



Online Assessment Guidelines (pages 127-132)

- District Summary – March 2014

Needs/Adequate	< 50%	50% – 80%	> 80%
Network	132 (44%)	8 (00%)	9 (00%)
Device	73 (27%)	83 (30%)	85 (29%)

- School Summary – March 2014

Category	Ready/Adequate	Needs	No Data
Network	3,656 (95%)	86 (2%)	55 (1%)
Device	3,094 (82%)	446 (12%)	207 (5%)

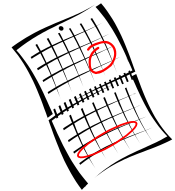
- Online Assessment Guidelines are based on technology readiness status of schools and districts that will be updated this spring.
- We expect that the vast majority of schools will be using online platform for both Part I and Part II.

TNReady Forms & Versions

- Multiple Forms
 - 3 - 4 forms per grade and subject. Up to 30 versions of an online form.
 - Only one paper form per grade and subject.
 - 30 common items, 20 unique items, 10 field test items?
 - Forms will be spiraled within a testing session
 - Post-equating process will ensure comparability between forms for scoring purposes
- What is **equating**? Procedure that allows for test scores to be compared across years as well as between forms
 - Pre-Equating** – To *predict* the relationship between the new and old test forms, ensure the similarity between the forms, and generate the proposed RS to SS tables
 - Post-Equating** – To *verify* the pre-equating outcomes and make sure that accurate RS to SS results are applied in final scoring

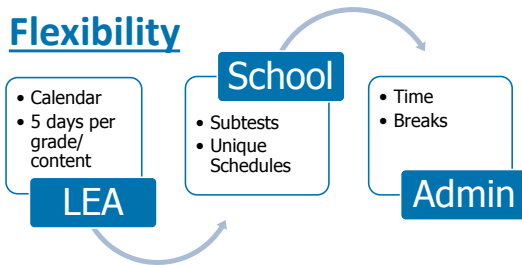
Administration Guidelines

- 2015 – 2016 TCAP Calendar (page 137)
- TNReady Draft Proposed Testing Windows (page 139)
 - Expected time per subtest:
 - Approximately 60-90 minutes
 - Exact time will be based on psychometrics
- TNReady Administration Guidelines (page 141)



Scheduling (page 141)

Flexibility



Scheduling (page 141)

Online



- Part I will only be offered online
- Flexible schedule
- Longer test window

Paper



- Part II will be available on paper
- Specific district-wide Schedule
- 4-day test window

Administration Procedures (page 141)

Create a secure, positive environment for testing...



AND.....

Reports & Timeline



Performance Levels

- We will continue to have 4 Performance Levels, as we do now.
- The Assessment Practices Task Force will discuss the names of each performance level and overarching performance level expectations at our June meeting.
- Teachers will inform the cut scores based on Performance Level Descriptors.
- Specific designation of which level will correspond to college and career readiness measure still to be determined.
 - UT and TBR will be reviewing the assessment frameworks against their remediation frameworks and will participate in determining the cut scores.

Timeline for TNReady Data Return Year One

- Part I and Part II will be reported as single score.
- Quick Scores – TBD
 - Standards setting will not be complete, so quick scores would have no relevance to PLDs.
 - Could not commit to quick scores being reported prior to June 2016.
- TVAAS – July 2016
- Achievement Reports
 - Student Reports – October 2016
 - School Report – October 2016
 - District Reports – October 2016

Communications



Key messages

- INSERT KEY MESSAGES FROM MICA VIDEO

Communications Discussion Questions

- INSERT HERE

Assessment Task Force Meeting 3:

Topics:

- Follow up conversation about principles
- Updates from SCORE about survey participation


To Do:

- Confirm parking
- Send prep
 - Colorado recommendations and other state
- Give ACT a heads up

Context 8:30-9:00	<ul style="list-style-type: none"> • Opening • Homework discussion (CM) 	<ul style="list-style-type: none"> • Bring copies of the minutes from last meeting
Quick Scores & Test Design and Development 9:00-9:30	<ul style="list-style-type: none"> • 5 minutes (Nakia and Candice) 	Prep
Principles 9:30-10:00	<ul style="list-style-type: none"> • 	
ACT 10:15-11:15	<ul style="list-style-type: none"> • Context • Legislation • What we are doing to support alignment to ACT • What are ACT/PLAN/EXPLORE/ASPIRE? (include cost and time) • What do we learn? What do we learn from other tests? • Value of the second test • Discussion • Recommendations 	<ul style="list-style-type: none"> • Legislation • Slides • Jonathan
K-2 11:15-12:15	<ul style="list-style-type: none"> • Context (Sat 10 going away) • Share the statute with mandatory K-2 • Desires for data in K-2, especially principal • RFP proposal pre-post 2nd grade • RTI – overview and screeners • Discussion • Other states and independent schools • Recommendations 	<ul style="list-style-type: none"> • PPT • Joey • RTI answer
Closing 12:15-12:30	<ul style="list-style-type: none"> • Add one more meeting 	

Next meeting:

- Accountability
- Opt Out
- Release Items
- Test Prep
- Overview and scoring and equating



Assessment Practices Task Force

Meeting Three
June 16, 2015

Task Force Goals

- Conduct an environmental scan of assessment usage and practices across the state
- Establish principles addressing purposes and goals of state assessments relative to locally chosen/designed assessments
- Define appropriate practices that best support decision-making at the state, district, school and teacher level
- Gain insight on ways to best communicate about TNReady to all stakeholder groups.

TENNESSEE DEPARTMENT OF
EDUCATION 1

Agenda

Opening & Discussion of Other States Reports	8:30-9:00
Quick Scores & Test Design Overview	9:00-9:30
Discussion of Principles from Meeting Two	9:30-10:00
ACT Discussion and Recommendations	10:15-11:15
Break	11:15-11:30
K-2 Discussion and Recommendations	11:30-12:25
Closing	12:25-12:30

TENNESSEE DEPARTMENT OF
EDUCATION 2

Other State Reports

- All state (Colorado, Iowa and Illinois) task forces/reports were required by legislation
- Colorado focused exclusively on principles and recommendations that were more specific in nature
- Iowa primarily focused on principles and ways to view testing generally
- Illinois' report focused primarily on the results of a statewide survey on testing
- Tennessee's report was not required by legislation
- Audience is legislators, district leadership and state department leadership
- Tennessee's report is focused on practices and includes principles, recommendations and a statewide survey


TENNESSEE DEPARTMENT OF
EDUCATION 3

Discussion About Other State Reports

- What are the primary areas of focus for other state's recommendations?
- What did you like about these reports and what was not helpful?
- What do you want to be sure our report includes?

TENNESSEE DEPARTMENT OF
EDUCATION 4

Quick Scores



TENNESSEE DEPARTMENT OF
EDUCATION 5

Test Development



6

Test Development



TENNESSEE DEPARTMENT OF
EDUCATION

7

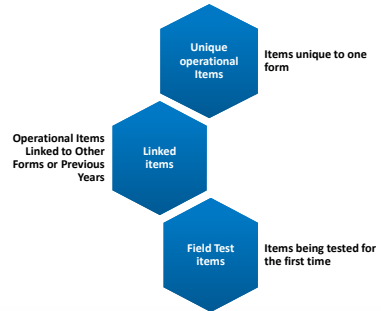
Test Design Terms

- Performance Level Descriptor: a description of the distinction between content mastery at different performance levels (basic, proficient, advanced)
- Blueprint: an out of the range of questions and point values for each category
- Field test: A trial run of an item that has been reviewed to test the performance of the item
- Form: A complete operational test
- Version: Different versions of a form can include different field test items

TENNESSEE DEPARTMENT OF
EDUCATION

8

Test Components



TENNESSEE DEPARTMENT OF
EDUCATION

9

Field Test Items

- Field test items get reviewed for:
- Student response to questions (% correct, question difficulty)
 - Disproportionate impact on specific student subgroups
 - Final content review

TENNESSEE DEPARTMENT OF
EDUCATION

10

Discussion of Principles



11

Discussion on Principles to Date

- Review notes from our last meeting
- Review draft of combined principles from last meeting
- Discuss what is missing or what needs more clarification in the areas of
 - Annual/Summative/Standardized Assessment
 - Interim/Benchmark/Formative/District Assessment
 - Test Preparation
 - Test Logistics

ACT



K-2



Legislative Context

- **49-6-6002. Tests not to be conducted earlier than grade three (3) -- Test dates.**
 - (a) No state-mandated test shall be conducted earlier than grade three (3), except that when the first and second grade tests provided for in chapter 434, § 7 of the Public Acts of 1997 are available, these tests shall be conducted.

K-2 Summative Context

- Optional assessment: SAT 10
- Purpose:
 - To measure student mastery in early grades
 - To inform intervention and remediation around numeracy and literacy
 - Can be used for value added for K-3 teachers (after baseline year)
- Norm referenced
- SAT10 will be phased out after 15-16

RTI² Context

- How have we handled intervention needs in TN prior to RTI²?
- Typically students were referred to special education when they struggled academically
 - Interventions were not aligned to areas of deficit
 - Operated under a "wait to fail" model

Response to Instruction and Intervention RTI²

GUIDING PRINCIPLES Leadership Culture of Collaboration Prevention & Early Intervention

TIER I - All
All students receive research-based, high-quality, general education instruction. In general, 80-85 percent of students will receive only Tier I instruction.

TIER II - Some
In addition to Tier I, extra help is provided to students who fall below the 20th percentile in basic math and reading skills. In general, 10-15 percent of students will receive Tier II interventions.

TIER III - Few
In addition to Tier I, extra help is provided to students who have not made significant progress in Tier II, are 2½-3 grade levels behind, or are below the 10th percentile in basic math and reading skills. Tier III interventions are more explicit and more intensive than Tier II interventions. In general, only 5-9 percent of students will receive Tier III interventions.

TENNESSEE DEPARTMENT OF EDUCATION
EDUCATION 18

Required Elements:

- Universal Screening
- Tier I
- Tier II and Tier III
- Progress Monitoring
- District and School RTI² Teams
- Fidelity of Implementation
- Parent Contact/Communication
- Highly trained personnel

TENNESSEE DEPARTMENT OF EDUCATION
EDUCATION

RTI² Universal Screeners

- Universal screeners are not assessments in the traditional sense. They are meant to be quick, informative, and non-intrusive and are given 3 times a year to identify at risk students.
- The most common Screeners selected and used by districts in grades K-5 are:
 - AimsWeb (30)
 - EasyCBM (15)
 - STAR 360 (47)
- In most of these programs, reading fluency is assessed for one minute with individual students while math computation is group administered and takes 5-8 minutes. This amounts to less than 30 minutes per child throughout the year.

TENNESSEE DEPARTMENT OF EDUCATION
EDUCATION

Guidance for Universal Screeners

- Skill-Based
- Nationally Normed
- Easy to Administer
- Limited Impact on Instructional Time
- Compares Apples to Apples (Alternate Assessment Reliability)
- Explicitly Measures each Skill Area
- Peer Reviewed in Special Education Journals

TENNESSEE DEPARTMENT OF EDUCATION
EDUCATION

Areas of Deficit for Intervention

- Basic reading skills (letters, letter sounds, phonological awareness, phonics)
- Reading comprehension
- Reading fluency
- Written expression
- Math calculation (column addition, basic facts, complex computation, decimals, fractions, conversions, percentages, etc.)
- Math reasoning/problem solving (number and operations, base ten, place value, measurement and length, fractions, geometry, algebra, expressions, linear equations etc.)

TENNESSEE DEPARTMENT OF EDUCATION
EDUCATION

How will RTI² benefit the students of TN?

- Early intervention to meet student need
- Problem solving process for each student
- Progress data to document student growth
- Closes achievement gaps
- Increases achievement results
- Better post-secondary outcomes and opportunities for students

TENNESSEE DEPARTMENT OF EDUCATION
EDUCATION 23

2nd Grade RFP

- Why 2nd grade?
 - Desires for data
 - 3rd grade remediation statute
 - Critical importance of strong foundations
 - Desire for primary school value add
- Design:
 - Pre-test, post-test
 - Criterion Referenced Test (based on standards)
 - Focused on reading and math

Discussion Questions

- What questions does this raise?
- What additional context would the group like to share about your perspective on K-2 assessment?

Principles & Recommendations

- What are some principles related to K-2 testing?
- What are our recommendations related to K-2 testing?

Closing



Notes

- Next meeting: Wednesday, July 15, 8:30-12:30
- Additional Meeting: Monday, August 24, 8:30-12:30
- Final Report: by mid-September



A Guide to Understanding Quick Scores

What vocabulary do I need to know?

- **Raw score:** the number of questions a student answered correctly.
- **Scale score:** generated from the raw score; this allows us to compare student performance from year to year. Because the questions on the test change each year, the level of difficulty could fluctuate. Scale scores give us a metric to make apples to apples comparisons from year to year.
- **Cut Score:** generated from the scale score; determines a student's performance level (i.e., below basic, basic, proficient, advanced).
- **Quick Score:** generated from the raw score; it is only meant to be factored into a student's end of year grade, as required by law. This score is on a 100-point scale because student grades are on a 100-point scale. Quick scores are not tied to student performance levels on TCAP (i.e., a quick score of 85 *does not* mean a student is proficient).

How are quick scores calculated?

Quick scores are generated from a student's raw score, or the number of questions they answered correctly. There are various methodologies that can be used to create a quick score from the raw score, and, this year, in an effort to use the same methodology for all grades, the department used the [cube root method](#) for grades 3-8. We have used the cube root method in high school since 2011.

What does the quick score mean?

A quick score creates a measure on a 100-point grading scale. It is *not* a percent correct or a percentile rank. Quick scores are generated only to be factored into a student's end of year grade. For grades 3-8, there is no statewide uniform grading policy. This means that a quick score of 84 in one district may be considered a *B*, while in another district it may be considered a *C*.

Has the bar for proficiency changed?

The bar for proficiency has not changed; it is the same as it was last year, and it is still based on the same scale scores. Because we know that the questions on the test change from year to year, sometimes questions may be slightly easier or more difficult. We generate scale scores to provide an apples-to-apples comparison. This is how we ensure that students are not penalized for the differences in test questions from year to year. So, in comparing this year to last year, students had the same expectations for performance in order to score proficient.

How do quick scores impact teachers?

Quick scores do not impact teacher evaluation or TVAAS, and quick scores have no bearing on personnel decisions tied to teacher performance.



A Guide to Understanding Quick Scores

How do quick scores impact students?

Because districts across the state do not have a uniform A-F grading policy for grades 3-8, the methodology we have always used to calculate quick score converts student performance on TCAP to the common 100-point grading scale.

For example, a student that answers 72 percent of questions correctly meets the bar for proficiency in third grade English language arts. However, if we were to use 72 percent as the student grade, in most districts that *proficient* student would receive a low pass or failing grade according to the typical A-F grading system. This is why quick scores are calculated on a 100-point grading scale; because TCAP performance levels are different from the A-F grading system.

How can my student have a quick score of 89 and still not be proficient on the TCAP?

A quick score does not determine TCAP performance levels (i.e., below basic, basic, proficient, advanced). Only scale scores, the comparable score generated based on the number of questions a student answered correctly, determine proficiency. For example, a quick score means different things in different subjects and grades. A quick score of 85 in seventh-grade science is equivalent to a scale score that is proficient, but a quick score of 85 in seventh-grade English language arts is equivalent to a scale score that is basic.

Quick scores are not intended to be a parent's primary window into their student's performance on TCAP. Quick scores are solely created for districts to incorporate student results into end of year grades. Detailed reports on student performance are shared with teachers and parents later in the summer to explain where students excelled and where they struggled.

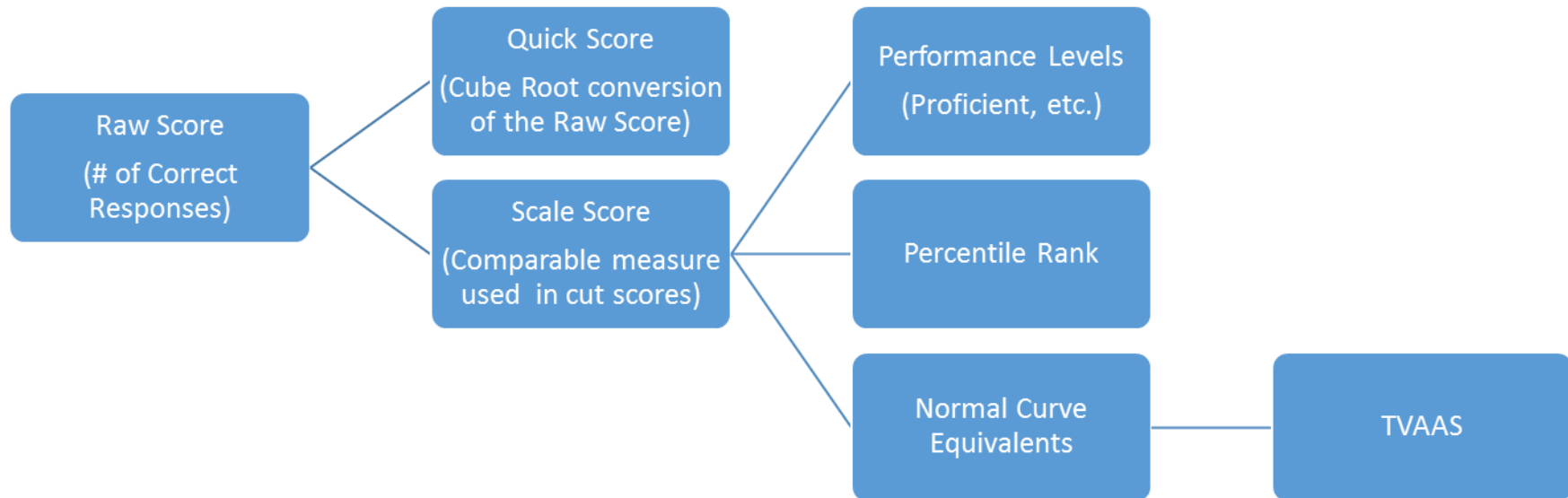
Why do we calculate quick scores if they aren't tied to proficiency levels?

State law requires that districts incorporate TCAP results into end of year student grades. In 2011-12 the department produced quick scores for grades 3-8 for the first time. Quick scores are not used for school or district accountability. They are not used for teacher evaluation or TVAAS.

As we transition to new assessments next year, we are revisiting the methodology for calculating quick scores in grades 3-8 given that there is no uniform grading policy across the state.



TCAP Scoring Flow Chart



Additional Information and Support on 2015 Quick Score Release

- **Monday, May 18**
 - Quick scores released to districts using cube root scaling methodology for grades 3-8.
- **Tuesday, May 19**
 - Commissioner learns of fall 2014 decision to change 3-8 quick score calculation and determines this change has not been previously communicated.
- **Wednesday, May 20**
 - Commissioner emails directors indicating quick score methodology change and timeline for additional information.
- **Friday, May 22**
 - Detailed information released to districts regarding the quick score methodology for grades 3-8 achievement and high school, as well as tables showing quick score that is equivalent to scale score proficiency cuts.
- **Wednesday, May 27**
 - Conference call with directors and district data and accountability personnel to answer questions.
- **Thursday, May 28**
 - Released raw scores for spring 2015 achievement 3-8 and high school to all districts. Second conference call with district data and accountability personnel.
- **Friday, May 29**
 - Launched *Understanding Quick Scores* website:
www.tn.gov/education/data/quick_scores.shtml
- **Friday, May 29**
 - Released *Educator Update* and *Principal Update* with message from commissioner and additional quick score information.
- **Friday, May 29**
 - Released *Classroom Chronicles* blog post from commissioner with apology for department communication failure and additional quick score information for all educators: <http://tnclassroomchronicles.org/clarifying-quick-scores/>
- **Saturday, May 30**
 - Released raw scores and corresponding quick score using interval-scaling methodology for spring 2015 achievement 3-8.
- **June 1-15**
 - Regional face-to-face meetings scheduled in East, Central and West Tennessee to answer additional questions for district personnel regarding quick scores and TCAP.
- **June 1-15**
 - Additional meetings with various stakeholder groups.



Cube Root Quick Score Formula and Calculation Procedures

Formula

$$\text{Cube Root Quick Score} = [(\text{Percent Score})^{(1/3)}] * 21.5443,$$

Where:

- Number Correct = Raw Score
- $(\text{Raw Score}/\text{Total Test Items}) * 100 = \text{"Percent Score"}$
- " $^{(1/3)}$ " is for cube root and
- "21.5443" is a constant term to make a maximum score equals to 100. The minimum score equals 0. Thus, the range of quick scores is from 0 – 100.

Note: We use four (4) decimal places for percent scores and cube root.

Procedures

Step 1: Calculate Percent Scores based on Number Correct

Step 2: Use Percent Score with four decimal places in the above formula.

Example 1

If raw score = 23 and the total number of test items = 50:

- then the percent score = 46.0000,
- the cube root of the percent score = $(46.0000^{(1/3)}) = 3.5830$, and
- the cube root quick score = $3.5830 * 21.5443 = 77.1932 = 77$.

Example 2

If raw score = 23 and the total number of test items = 64:

- then the percent score = 35.9375,
- the cube root of the percent score = $(35.9375^{(1/3)}) = 3.3000$, and
- the cube root quick score = $3.3000 * 21.5443 = 72.0962 = 72$.



Charting a Path Forward on ACT, PLAN, and EXPLORE

Assessment Practices Task Force
June 2015

TCA §49-6-6001 requires Tennessee students to take standardized exams in 8th, 10th, 11th grades

As a strategy for assessing student readiness to enter and succeed in postsecondary training, every public school student shall take a series of three examinations, one administered at grade eight, one administered at grade ten, and one at grade eleven. These assessments shall be approved by the commissioner of education and provide educators with diagnostic information to assist in developing interventions for the purpose of increasing high school graduation rates and improving student preparation for postsecondary achievement.

ACT scores play a major role in shaping Tennessee students' future opportunities

- ★ A score of 21 or higher on the ACT makes students eligible for the Tennessee HOPE scholarship
- ★ ACT scores are used by most Tennessee postsecondary institutions to identify students in need of remediation (generally students scoring below 19)
- ★ Many industries use ACT scores as a selection mechanism for entry-level jobs

TDOE has set ambitious goals for student success in high school and beyond

Success After Graduation

GOAL #2



The average ACT score in Tennessee will be a 21, allowing more students to earn HOPE scholarships.

GOAL #3



A majority of high school graduates will go on to earn a certificate, diploma, or degree.

Tennessee graduating cohort ACT scores have ranged between 19.0 and 19.3 over the last 4 years



Task Force Decision Point

How can we create a testing framework that offers sufficient information to students, schools, and districts to prepare for the ACT while minimizing the overall test burden?

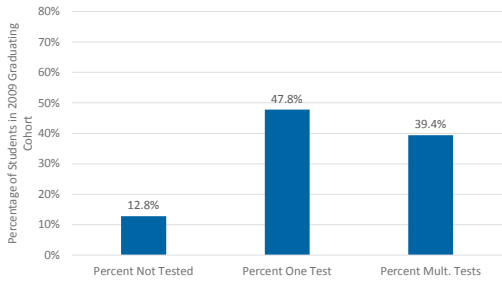
Points to consider...

- ACT tests and retests
- TNReady
- 8th Grade EXPLORE
- 10th Grade PLAN
- Revamped ACT Aspire

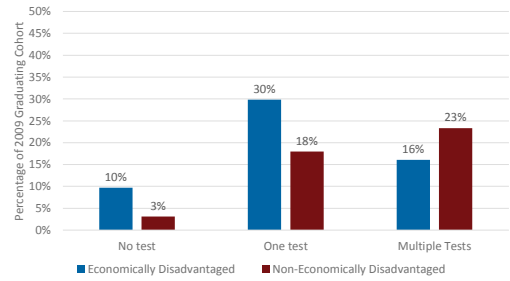
The Landscape of ACT Testing



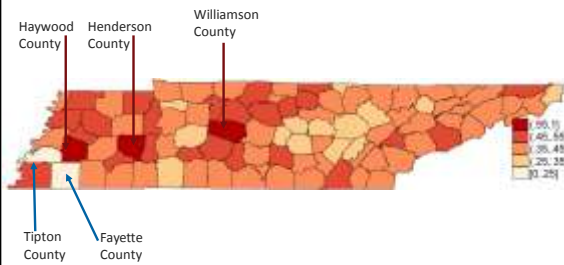
Around 40 percent of students took the ACT multiple times among 2013 graduates



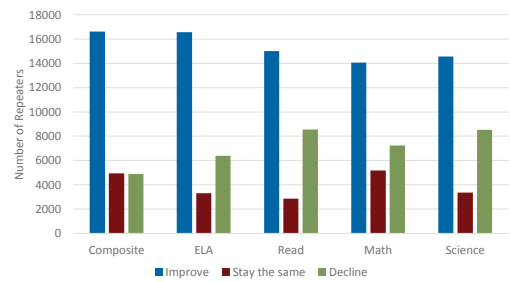
Economically disadvantaged students are almost 50 percent less likely to sit for the ACT on multiple occasions



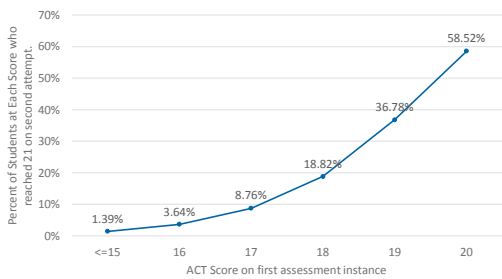
Not only do we see variation by socioeconomic status, but we see significant variation across counties



Student ACT scores generally increase when they retake after the first assessment instance



Retaking is particularly important for those who fall just below a 21. Half the students who scored a 19 or 20 score a 21 the second time they take the assessment.



# of Repeat Test Takers (%)	5,210 (28%)	1,998 (43%)	2,128 (48%)	2,151 (49%)	2,152 (51%)	2,103 (55%)
-----------------------------	-------------	-------------	-------------	-------------	-------------	-------------

Testing in 8th and 10th grade



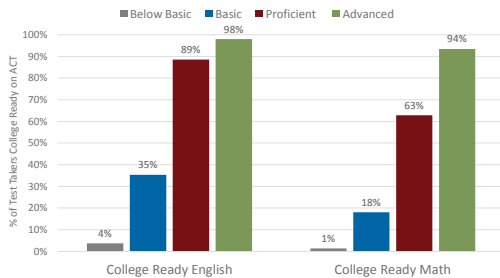
EXPLORE and PLAN will be phased out by ACT next year and replaced by the Aspire exam

- ★ EXPLORE
 - 120 minute test in 8th grade
 - Cost: \$10.15 per pupil
- ★ PLAN
 - 115 minute test in 10th grade
 - Cost: \$10.50 per pupil
- ★ As a state, we spend roughly \$1.3 million per year on the PLAN and EXPLORE
- ★ ASPIRE
 - Approximately 4 hours for middle and early high school
 - Cost: 14.70 per pupil online or \$20.70 per pupil on paper

The 8th grade EXPLORE and 10th grade PLAN correlate strongly with ACT performance

- ★ In English...
 - 79% of students who are "college ready" on 8th grade EXPLORE score at a college ready level on the ACT
 - 83% of students who are NOT "college ready" on 8th grade are still not college ready on the ACT
- ★ In Math...
 - 55% of students who are "college ready" on 8th grade EXPLORE score at a college ready level on the ACT
 - 97% of students who are NOT "college ready" on 8th grade are still not college ready on the ACT

But our 8th grade TCAP tests also tend to be highly predictive of later ACT performance



Task Force Discussion

- ★ Can you provide other context from your experience that might inform the state decision around TNReady, Aspire, EXPLORE, and PLAN?
- ★ What are your perceptions of PLAN and EXPLORE?
- ★ How can we as a state create greater equity around ACT retests?
- ★ Are there specific principles and/or recommendations related to ACT that should be set by the task force?

Assessment Practices Task Force: Meeting Four - July 15, 2015

Task Force Goals:

- To conduct an environmental scan of assessment usage and practices across the state.
- To establish principles addressing purposes and goals of state assessment relative to locally designed or locally chosen assessments (i.e. formative assessments).
- To define appropriate practices associated with these principles that best support decision-making at the state, district, school and teachers levels.
- To gain insight on ways to best communicate about TN Ready to all stakeholder groups.

Meeting Four Goals:

- Discuss findings from surveys, focus groups, and interviews about assessment
- Review current accountability system and information about proposed changes
- Review and provide feedback on draft principles and recommendations for state and local assessments

Agenda:

Topic	Components	Timing
Welcome	<ul style="list-style-type: none"> ▪ Review the agenda and goals ▪ Review of progress ▪ 2015 TCAP results 	8:30-9:00
Overview of SCORE findings	<ul style="list-style-type: none"> ▪ Update on focus group work related to assessment/testing study 	9:00-9:20
2015 Educator Survey Results	<ul style="list-style-type: none"> ▪ Preliminary overall findings ▪ Discussion of assessment specific survey items and results 	9:20-9:45
Break		9:45-10:00
Accountability	<ul style="list-style-type: none"> ▪ Review information about district, school, and teacher accountability system ▪ Learn about proposed changes 	10:00-10:45
Review draft of principles and recommendations	<ul style="list-style-type: none"> ▪ Review draft ▪ Discussion of draft principles ▪ Discussion of draft recommendations ▪ Identify additional principles or recommendations to include 	10:45-12:00
Closing	<ul style="list-style-type: none"> ▪ Whole group discussion ▪ Next steps 	12:00-12:30

Future Meeting Date: August 24, 2015, 8:30-12:30

Groups for Discussion:

Group 1	Group 2	Group 3
Candice McQueen	Nakia Towns	Vicki Kirk
Sara Heyburn	Dolores Gresham	Nicole Roberts
John Forgety	Wanda Shelton	Harry Brooks
Philip Eller	Bill O'Donnell	Beth Unfried
Jasmine Carlisle	Stephen Smith	Debbie Shedden
Danielle Mezera	Audrey Shores	Barbara Gray
		Kathleen Airhart



Assessment Practices Task Force
Meeting Four
July 15, 2015

Task Force Goals

- Conduct an environmental scan of assessment usage and practices across the state
- Establish principles addressing purposes and goals of state assessments relative to locally chosen/designed assessments
- Define appropriate practices that best support decision-making at the state, district, school and teacher level
- Gain insight on ways to best communicate about TNReady to all stakeholder groups.

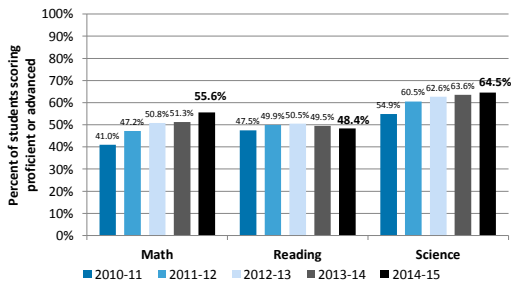
Agenda

Time	Topics
8:30-9:00	Welcome
9:00-9:20	Overview of SCORE findings
9:20-9:45	2015 Educator Survey Results
9:45-10:00	Break
10:00-10:45	Accountability Overview
10:45-12:00	Review of Draft Principles and Recommendations
12:00-12:30	Closing

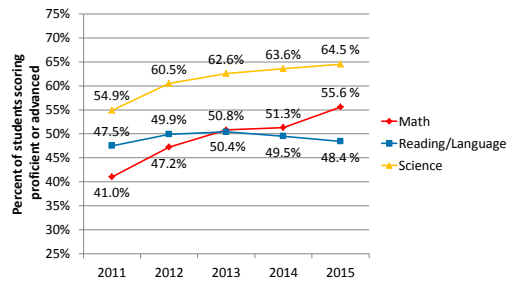
All Students



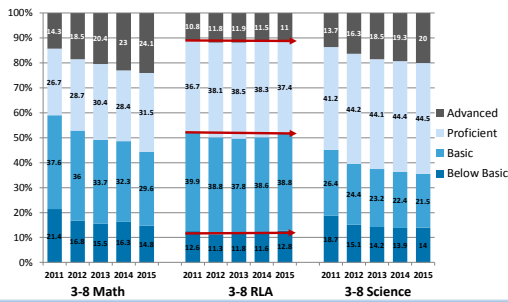
In grades 3-8, students continued to make gains in math and science, with a slight decrease in RLA.



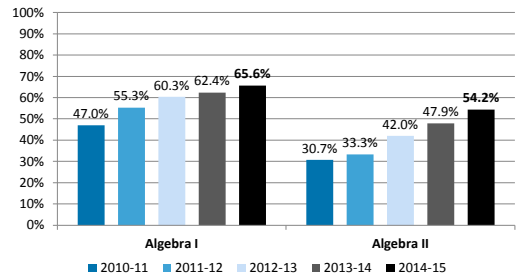
The rate of student progress in math has eclipsed results in reading over the past five years.



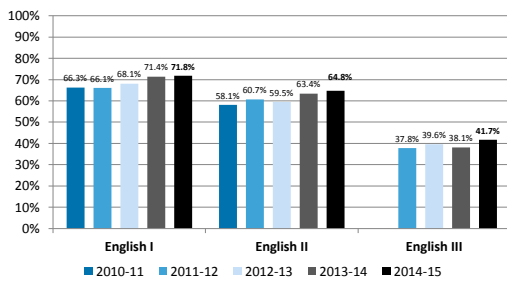
Not only is 3-8 RLA proficiency lagging other content areas, the distribution across performance levels has also remained steady in RLA.



High school students continued to make large gains in math.



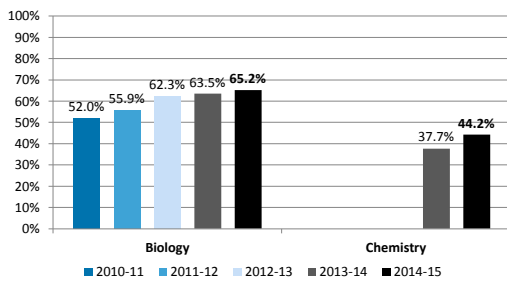
Students also improved in high school English, with the largest increase in English III.



Questions

- Do the current recommendations adequately reflect the principles?
- What is missing or unclear?
- What has not been discussed that you would still like to discuss or consider?
- How do you think these principles and/or recommendations will change practice?
- What can we do to share, promote, and align practice to these principles and/or recommendations?

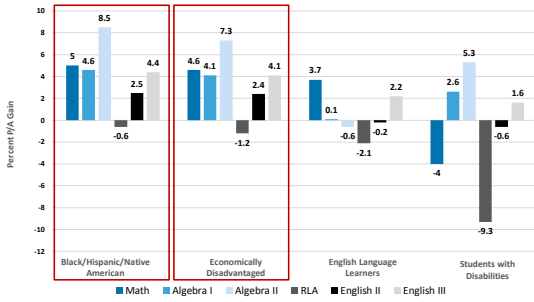
Similarly, students made progress in high school science, with Chemistry results greatly improving over last year.



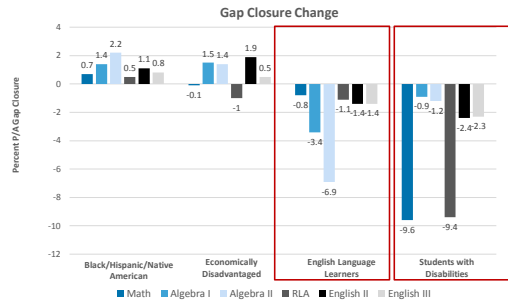
Subgroups and Gaps



BHN and ED students had consistent and substantial gains in proficiency across subjects.

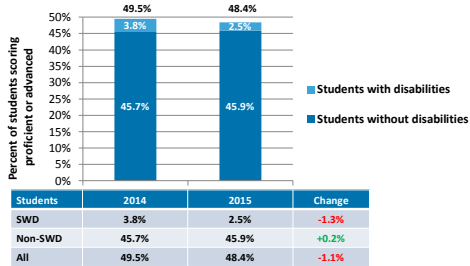


Gaps continue to widen between SWD and ELL students versus their comparison groups.



This year, we raised expectations for *Students with Disabilities (SWD)* by transitioning all non-portfolio students from MAAS to standard TCAP Achievement tests in grades 3-8.

These results reflect a new baseline for SWD performance.



State Accountability Results Summary

- Students made gains in 13 of 18 tests in 3-8 math, reading, and science.
 - 3-8 math and science increased.
 - 3-8 reading declined slightly, continuing a trend of comparatively weak results over the last four years.
- Students made gains in all high school subjects.
- Students made gains in math across all individual grades.
- Black/Hispanic/Native American (BHN) and Economically Disadvantaged (ED) student subgroups continued to show progress.
 - BHN students narrowed gaps in both 3-8 math and reading.
 - BHN and ED student subgroups narrowed gaps in all high school subjects.
- Achievement gaps for Students with Disabilities (SWD) and English Language Learners (ELL) increased.

Groups for Discussion:

Group 1	Group 2	Group 3
Candice McQueen	Nakia Towns	Vicki Kirk
Sara Heyburn	Dolores Gresham	Nicole Roberts
John Forgety	Wanda Shelton	Harry Brooks
Philip Eller	Bill O'Donnell	Beth Unfried
Jasmine Carlisle	Stephen Smith	Debbie Shedden
Danielle Mezera	Audrey Shores	Barbara Gray
		Kathleen Airhart

SCORE Assessment Study

July 15, 2015

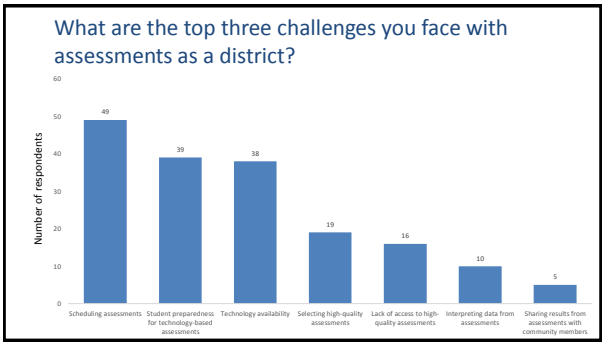
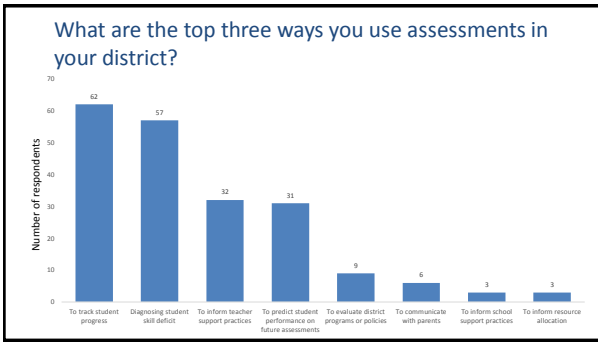
Overview

- Update on SCORE study
- Focus group findings
- District interview findings
- Q&A

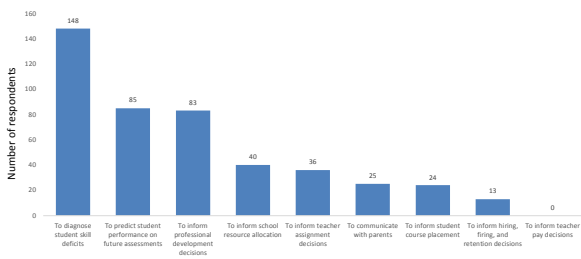
Updates

- Completed focus groups with principals and teachers (April-June 2015)
- Completed district interviews (June 2015)
- Data analysis (June-July 2015)

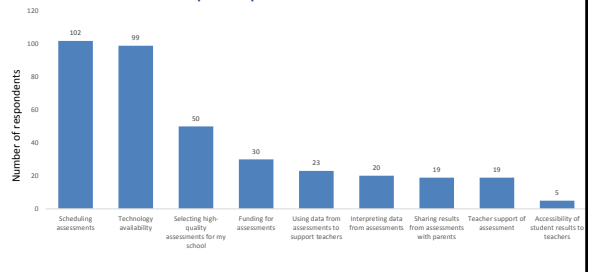
Recap: Survey Results



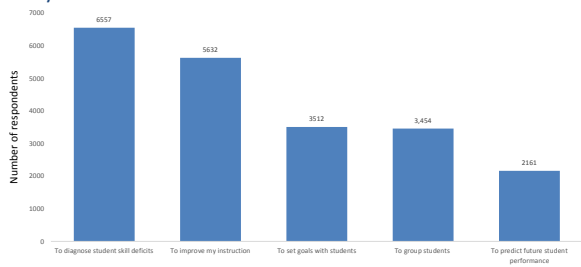
What are the top three ways you use assessment as a principal?



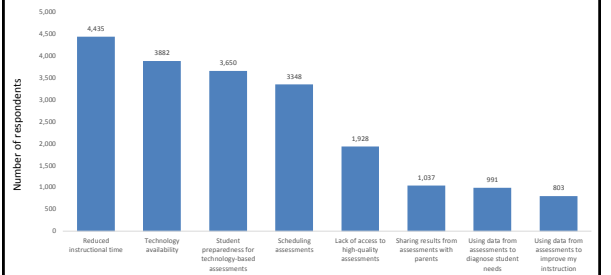
What are the top three challenges you face with assessment as a principal?



What are the top three ways you use assessment in your classroom?



What are the top three challenges you face with assessments as a teacher?



Focus group sample

- Conducted nearly 40 focus groups in 8 different communities
- Met with about 220 teachers, principals, and central office staff from districts across the state

Focus Group Results

Focus group questions

- Describe the use of interim or benchmark assessments in your school and district. Are these assessments useful to you? If so, why? If not, why not?
- How would you describe the amount of time spent on assessments in your school?
- What are the top challenges you face with assessments in your school?

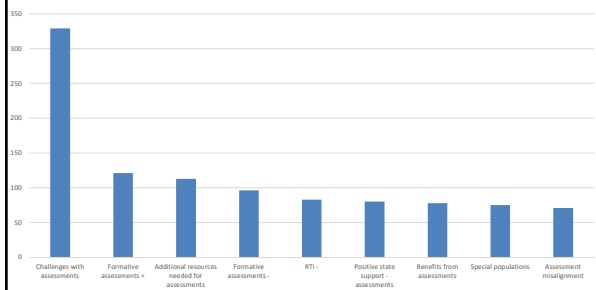


Focus group questions cont.

- Describe how you feel about Tennessee's transition to the TNReady assessment.
- What support do you need to ensure a successful transition to TNReady?
- Describe your district's readiness for the TNReady assessment from a technological standpoint.



Top codes by sub-theme



Challenges with assessment

- Time spent on assessment
- Challenges with technology-based assessments
- Limitations of current assessments
- Special populations
- Student anxiety



"You have to lay out a calendar. We just had a principal's meeting in our district and we looked at our calendar. We highlighted with colors the different tests and whether it's no teaching that day. I thought, what if you give teachers this calendar and say, 'Go find your days to teach.' Those are important."


-Teacher



"But do you want to see a child when they're truly at their happiest? You have to go to a PBL, a project-based learning experience, where they've got their hands dirty, they're making, they're thinking, they're collaborating, they're creating something fantastic, they're pumped. Okay, have you ever walked into a TCAP and seen somebody pumped? Maybe there's a few that really get excited about that. But if you really want somebody to be prepared for the workforce, you get them collaborating, you get them in there thinking and creating and making something. We didn't go to the moon taking a TCAP. We go and we create and we build as a team. That's where we build our workforce. But we're missing that and I just think it's somewhat more simple than we're making it, but we're making it a bureaucracy and confusing and it doesn't have to be that way."

-Principal






Formative assessments +/ Formative assessments -

- Benefits from formative assessments
- Teacher created common formative assessments
- Limitations of current assessments
- Misalignment


© 2014 Pearson Education, Inc. All rights reserved. E-CORE



“I think formative and benchmark assessments are critical and vital for teachers in the elementary school to understand the progress the students are making. But once again, if they’re excessive it takes away instructional time as has been mentioned. But, they’re extremely important in the elementary division, but it’s the way they’re utilized that makes the difference.”

–Teacher

© 2014 Pearson Education, Inc. All rights reserved. E-CORE



“Well, they are huge. I spend lots of time looking at where we’re going, how quickly are we going, are we moving kids? Not only are we moving groups of kids, are we moving individual students? We’ve completed our last benchmark and I’m real anxious to get it back. How many of my RTI kids am I moving? It just leads us and it drives us. I believe that it’s a wonderful tool.”

–Central Office


© 2014 Pearson Education, Inc. All rights reserved. E-CORE



Additional resources

- Funding/financial limitations
- Technology support
- Statewide benchmark assessment/state guidance
- TNReady
- RTI²

© 2014 Pearson Education, Inc. All rights reserved. E-CORE



“I think that you need to set teachers and students up for success and not for failure. It doesn’t need to be a game of trickery where we’re guessing. It just needs to be clear. They need to be open about it. And say, ‘This is exactly what it’s going to look like.’ And I know that’s a lot of work. I know that’s a lot to get out that soon, but I agree. It needs to be here in August so that you can hit the ground running.”

–Teachers

© 2014 Pearson Education, Inc. All rights reserved. E-CORE



“We’ve seen some successes with RTI. The only drawback with it that I see is not having enough personnel in the school to really carry it out the way that I think it is designed to be. It’s a really good idea to have one teacher to three kids, but when you start to put that into reality, it doesn’t look like that.”

–Principals

© 2014 Pearson Education, Inc. All rights reserved. E-CORE

“What would my ideal assessment look like? Something that looks like the test at the end. Like she said, the state produces that. Why can't they produce some assessments that we could use all winter and spring, that's going to replicate the TCAP test? Instead of us grabbing something over here or something over there, I would like to see something like that. I think that's a pretty good idea that the state could do that. They live with the TCAP test, so generate some [benchmark] assessments.”

-Teacher



District Interview Results



District interview questions

- How many interim or benchmark assessments are administered in your district? About how many days does it take to administer these assessments?
- How do you use assessments in your district? In what ways do you use data from assessments to drive decisions in your district?
- How do you select benchmark or formative assessments to implement in your district? Do you have any processes in place that you use to select assessments? If so, could you walk us through that process?

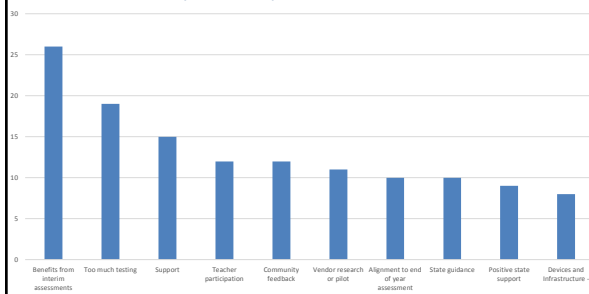


District interview questions cont.

- Describe how you feel about Tennessee's transition to the TNReady assessment.
- How has the state communicated with you about the changes to Tennessee's assessment?
- What support do you need to ensure a successful transition to TNReady?
- Describe your district's readiness for the TNReady assessment from a technological standpoint.



Top codes by sub-theme



Benefits from interim assessments

- District use of interim assessment data
- Teacher use of interim assessment data
- Program evaluation tool



"It does influence next steps. If a child is responding and the trend is beginning to increase. Then you know, this is working, I'm going to continue this. When I check in next week or in two weeks, I want to see what's happened now. If a child is flat-lining, then we know that what we're currently doing doesn't seem to be having an impact. We ask, what changes can I make to make a difference? He's in a group of five, maybe I need to put him in a group of three. He only gets 3 turns in a 30 minute cycle, maybe I need to double his turns. Maybe I need to change the interventionist, maybe I need to change the time of day. So there's all these variables that need to be looked at, but you have to use the data to drive that decision."

-District leader



Too much testing

- Teacher feedback
- Principal feedback
- Parent and community feedback
- Challenges faced in efforts to reduce the amount of assessment in districts



"I think a lot of educational days, days that we should be spending teaching, we're spending testing. Whether it's benchmarking for RTI, practicing the new TNReady platform, or actually taking the end of course exam. I just think we spend a lot of time preparing for tests and testing, and that's time we could spend educating."

-District Leader



"When I think too much assessment, I'm thinking of hours spent arduously working on very challenging things and I don't really think that's what happening. I just think there's this misperception out there and it just feels like a wave that sort of took over. I'm not certain that's an actual reflection of what it looks like, but it's very hard to combat because I think there are people who really believe that we just assess too much, and it's hard to combat that."

-District Leader



Support

- Financial support/funding
- RTI²
- Technology
- State guidance/statewide benchmark assessment
- TNReady



"We could actually use a lot of support when we're talking about RTI universal screeners. In fact, it's state mandated that we are doing that, but there's no funding tied to that. That does create a hardship for school systems when you're mandated to do something, but yet there's not any type of funding tied to that."

-District Leader



"I know that the MICA system is there and they're going to load more items into it. But looking to the future, it would be good to have a more comprehensive bank where we can build our own formative, daily, weekly assessments that are right in line with the end of year assessment. It would probably line all these things up and be better data because the reporting would look alike, the items would look alike. I just think that would be a good tool."

-District Leader



Q&A





Tennessee Educator Survey

2015 Survey Preliminary Results

Survey Context

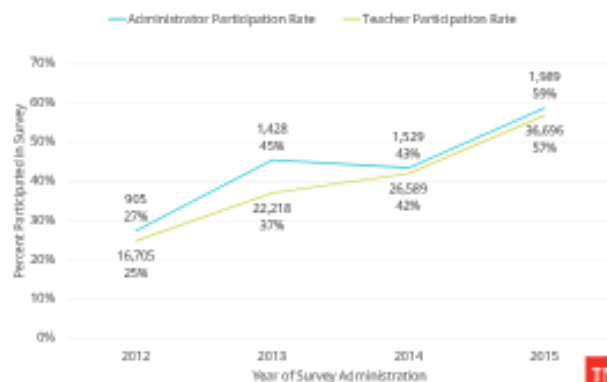
- The 2015 Tennessee Educator Survey was emailed to all teachers and administrators in Tennessee on April 16, 2015
- The survey window remained open through May 29, 2015.
- Surveys were differentiated by role
 - Teacher Survey
 - Building Administrator Survey
 - Certified School-Level Support Staff Survey
 - District Survey
- All administrators and teachers received a core survey and were randomly assigned one module that began after the completion of the core survey.

Survey Modules

- Data and Digital
- Teacher Evaluation
- Standards
- Professional Learning
- Assessment (Teachers Only)



Over 10,000 more teachers and almost 500 more administrators completed the survey this year



The Tennessee Educator Survey is designed to remain consistent for the next several years

- Purposes include:
 - State-Level:
 - Research and agenda-setting
 - Progress monitoring
 - Communications and talking points
 - Listening more to educators
 - District/School-Level:
 - District and school improvement planning
 - Reducing survey fatigue by streamlining TDOE survey process



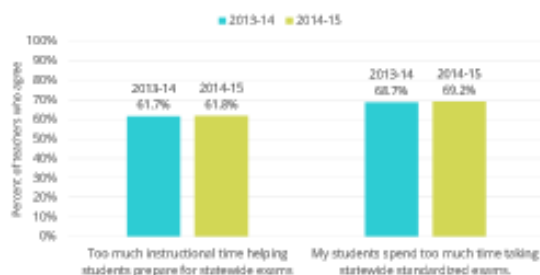
In August, a user-friendly site will publicly report survey results

- The site will...
 - Publish information that includes aggregated results at state level
 - Publish district- and school-level results for districts and schools with 50 percent participation and at least 10 potential respondents.
 - Almost 1,100 schools (66 percent of eligible schools) and 125 (84 percent of eligible districts) will receive teacher results, while almost 75 districts (77 percent of eligible districts) will receive administrator reports.
 - Data will be presented in a visually appealing manner with raw data available to download



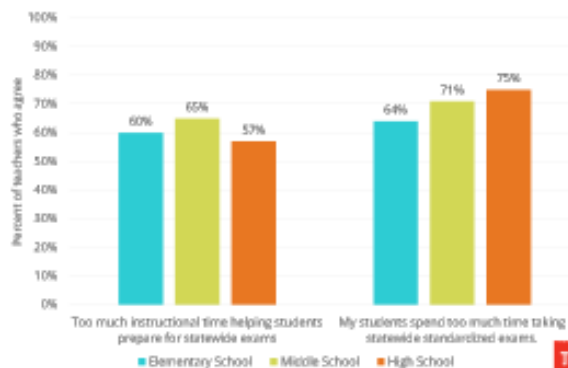
Time on Assessment

Most teachers feel like they are spending too much time on exam preparation and testing



TN

Middle school teachers are most likely to say they spend too much time preparing, but HS teachers are most concerned about testing time



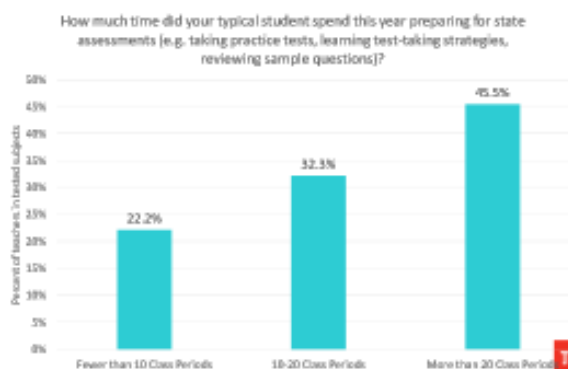
TN

Three quarters of teachers in tested subjects are concerned with the amount of time spent on statewide tests



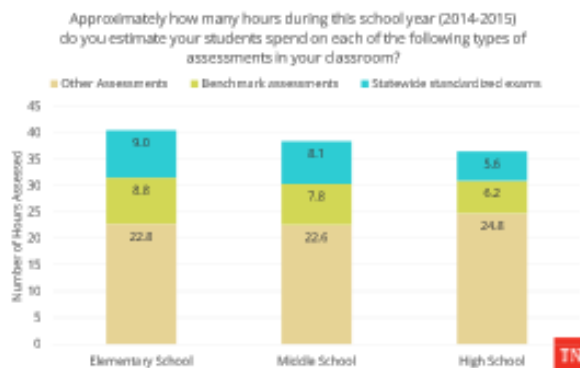
TN

The average student spends at least 20 class periods over the year in preparation for state assessments



TN

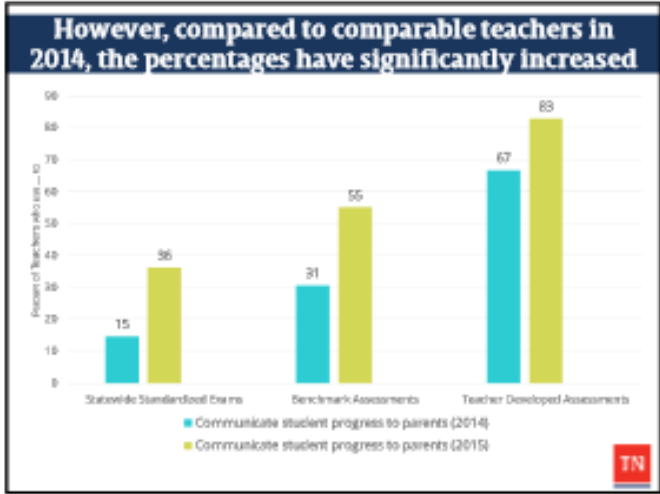
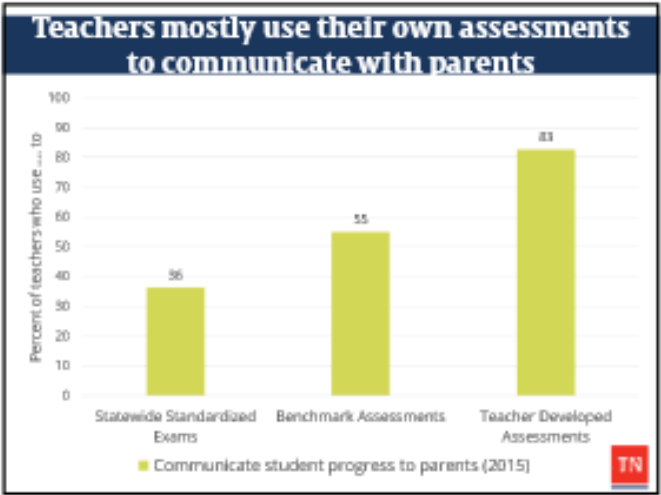
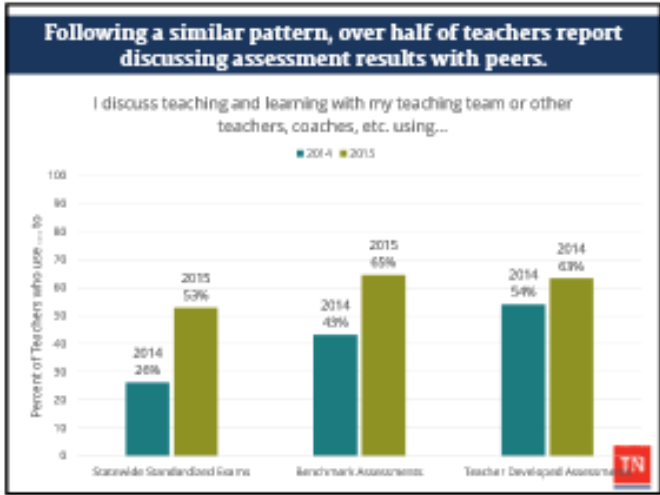
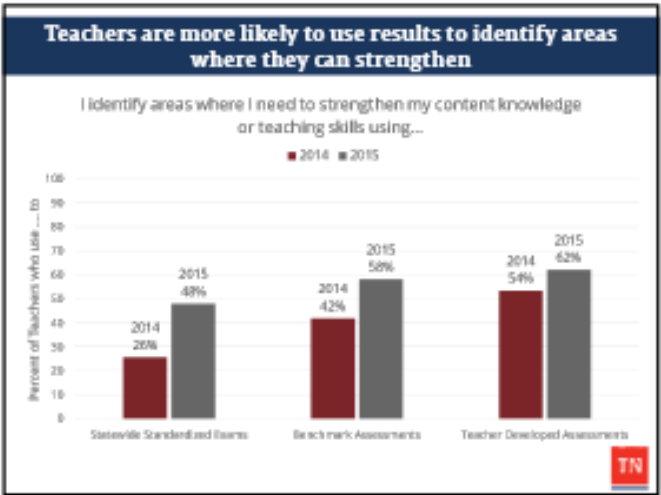
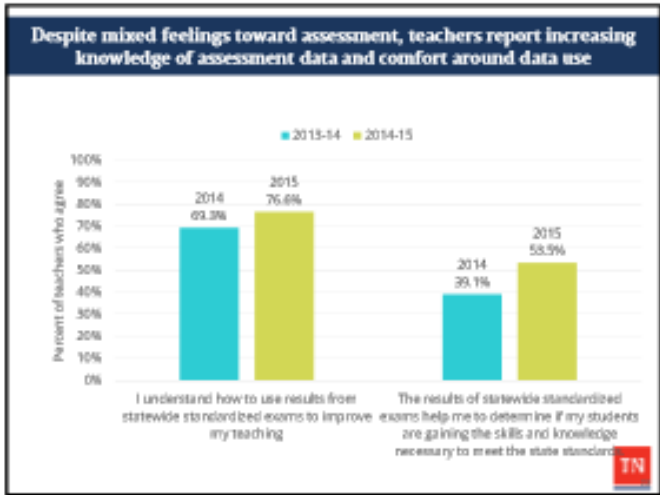
On average, elementary school teachers believe their students spend more time taking assessments than middle or high school teachers



TN



Use of Assessment Results





Accountability Overview

Assessment Practices Taskforce

Data & Research Division, July 15, 2015

Guiding Questions

- Why do we have **accountability systems**? What is the history?
- What accountability structures does Tennessee have for **schools and districts**?
- What changes are forthcoming with the **ESEA Waiver** for district and school accountability?
- What accountability structures does Tennessee have for **teachers and administrators**?
- What changes are forthcoming for teacher accountability with the **transition to TNReady**?



2



Accountability: Theory of Action

3

Why do we have accountability systems?

Typically, arguments for greater accountability in education come from two overarching sources: (1) Political (2) Theoretical

Political

- **Return on Investment** - Taxpayers should be able to measure and know the relative benefit of publicly funded endeavors.
 - Requires public reporting of performance metrics and progress against goals
- **Economic Development** - Higher levels of educational attainment may be reflected in improved labor productivity and innovation.
 - For individuals, the greater skills one has, the larger wages he can demand
- **Equity** - Overarching goal of serving all racial, ethnic, economic, or ability groups based on high expectations.
 - Disaggregating data through accountability frameworks to monitor equity goals



4

Theoretical: Principal-Agent Challenge

- **Agency relationship:** whenever one individual depends upon the action of another.
 - State and district goals for education are ultimately dependent upon the actions of administrators, teachers, and other educators.
 - Educators must motivate students to engage in meeting learning goals.
- Sometimes principals and agents may have different goals or priorities. One example:
 - State legislature (principal) wants to increase economic development through focus on STEM (science, technology, engineering, and math) education
 - Schools and teachers (agents) want to maximize social-emotional learning through co-curricular activities
- These differences may manifest into "agency loss" for organizations.
 - It is impossible to have "perfect information" on the daily activities of 80,000 teachers and 1,000,000 students toward goals



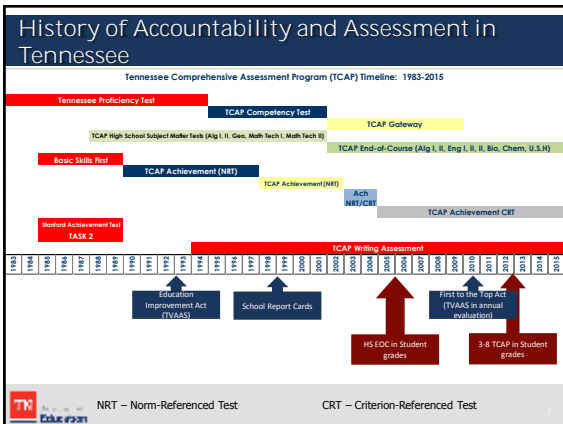
5

Theoretical: Principal-Agent Challenge

- **Monitoring**
 - Generally focused on outcomes - achievement test scores, graduation rates, attendance, etc.
 - Performance management including evaluation systems
- **Incentives**
 - Pay and promotion based on performance (as measured by outcomes or behaviors)
 - Recognition such as Reward schools, Exemplary districts, School Report Card
 - Greater autonomy, additional funding, or other resources
- **Sanctions**
 - Negative public determinations
 - Additional reporting, planning, and compliance obligations
 - Demotion or termination



6



Current accountability framework for districts, schools, and educators

District Determinations	School Lists	Educator Evaluation
<p>Districts receive annual determination based on student growth in multiple areas including:</p> <ul style="list-style-type: none"> Achievement Tests in grades 3-8 End-of-Course Exams in high school courses High School Graduation 	<ul style="list-style-type: none"> Reward schools: Top ten percent of schools based on absolute performance and value-added growth Focus schools: Ten percent of schools with the largest achievement gaps Priority schools: Bottom five percent of schools, based on absolute performance 	<ul style="list-style-type: none"> All school-based personnel, including administrators and support services participate in annual evaluation. Multiple measures including: <ul style="list-style-type: none"> Observation rubrics Individual TVAAS or Portfolio Growth Measure Other achievement data (AP, Graduation Rate, etc.) Student survey (optional) Teacher survey (evidence for administrator rubric)

For the last 3 years, TN has been approved for a flexibility waiver that allows a state-driven accountability system

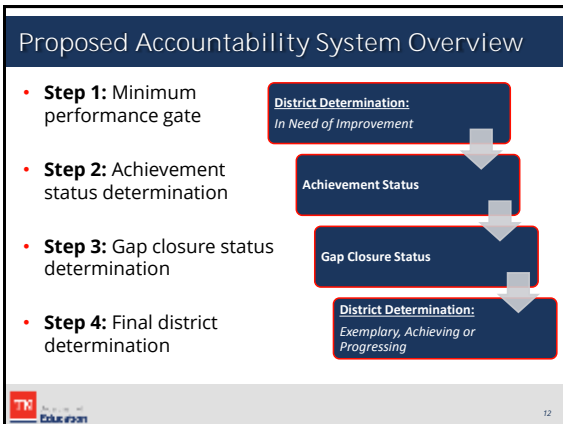
2011-12

- Tennessee earned approval for its first ESEA flexibility waiver from the federal department of education, developing its own state accountability system and
- Avoided having more than 75% of schools classified as failing under No Child Left Behind due to goal of 100% proficiency.

2014-15

- Tennessee achievement scores continue to improve – and the state prepares for the new TNReady assessment
- The state's ESEA flexibility waiver expires at the end of summer 2015
- Renewal Application submitted for approval March 31.

- ### The system proposed in waiver application...
- ...recognizes the hard work districts do to make *incremental gains* by rewarding partial credit for improving but not meeting targets.
 - ...recognizes districts that *greatly* exceed their targets or expected growth/performance.
 - ...will work for all years moving forward, with certain elements phased in as data become available.
 - ...includes many pathways to Exemplary.
 - Proposed District determinations:
 - Exemplary, Achieving, Progressing, and In Need of Improvement.



Key Changes in Proposed System

- **Grades 3-8 separated into Grades 3-5 and Grades 6-8**
 - Aligns cohort sizes at elementary, middle, and HS level
 - Offers districts more opportunities to show improvement (and a lower chance of failing the initial minimum performance gate)
 - Eliminates AMO targets for individual grades 3 and 7.
- **All HS EOCs included in HS English and HS Math**
 - Helps capture students who accelerate in middle school
- **ACT included as a measure in the system**
 - Ensures that accountability system reflects reality of education landscape and TDOE strategic priorities
- **Subgroups combined into super subgroup (BHN, ED, SWD, and ELL) at minimum performance gate**
 - No subgroup improvement test
 - Assesses performance of neediest students without disadvantaging more diverse districts



Key Changes in Proposed System

- **Achievement focus maintained**
 - Three pathways to demonstrate progress
 - Elevates TVAAS as a pathway, versus a safe harbor, and reflects a student-level cohort growth measure
- **Gap Closure focus maintained and calculation revised**
 - Eliminates Comparison/All - Subgroup = GAP; measured via subgroup performance via three pathways
 - Includes BB → B as pathway
 - Includes district subgroup TVAAS as pathway
- **Includes a performance scale rather than met/miss targets**
 - Beyond minimum performance gate
- **Eliminates *In Need of Subgroup Improvement* and *Intermediate final district determination***
 - Achievement and gap closure equally weighted in final determination



Current accountability framework for districts, schools, and educators

District Determinations

Districts receive **annual determination** based on **student growth** in multiple areas including:

- Achievement Tests in grades 3-8
- End-of-Course Exams in high school courses
- High School Graduation

School Lists

- **Reward schools:** Top ten percent of schools based on absolute performance and value-added growth
- **Focus schools:** Ten percent of schools with the largest achievement gaps
- **Priority schools:** Bottom five percent of schools, based on absolute performance

Educator Evaluation

- **All school-based personnel, including administrators and support services participate in annual evaluation.**
- **Multiple measures including:**
 - Observation rubrics
 - Individual TVAAS or Portfolio Growth Measure
 - Other achievement data (AP, Graduation Rate, etc.)
 - Student survey (optional)
 - Teacher survey (evidence for administrator rubric)



School Accountability

- **Reward schools designated annually**
 - **Performance:** Top 5% of schools in the state based on achievement (excludes schools where achievement gaps exceed the state)
 - **Progress:** Top 5% of school in the state based on TVAAS growth
- **Focus Schools**
 - Largest achievement gaps between groups in the state OR
 - Low performance for subgroup based on achievement or graduation rate
- **Priority Schools**
 - Lowest performing 5% of schools based on achievement (reading, math, and science)
 - Schools can be exempt if meeting annual goals
 - Priority schools may be eligible for inclusion in the Achievement School District



Achievement School District

- ASD was created as a **state-run school district** to turnaround schools in the lowest 5% (priority schools) of schools based on achievement of schools in Tennessee
- ASD has **direct-run schools** and also **authorizes charters** to serve as operators of priority schools
- ASD focuses on **attracting** high-quality, proven charters and supporting local charters to run schools
- ASD is **one strategy** to turn-around lowest performing schools.
 - Other strategies endorsed by the state include district-run I-Zone schools, state-funded SIG (School Improvement Grant) program, and monitored, progressive district plans for priority schools



Current accountability framework for districts, schools, and educators

District Determinations

Districts receive **annual determination** based on **student growth** in multiple areas including:

- Achievement Tests in grades 3-8
- End-of-Course Exams in high school courses
- High School Graduation

School Lists

- **Reward schools:** Top ten percent of schools based on absolute performance and value-added growth
- **Focus schools:** Ten percent of schools with the largest achievement gaps
- **Priority schools:** Bottom five percent of schools, based on absolute performance

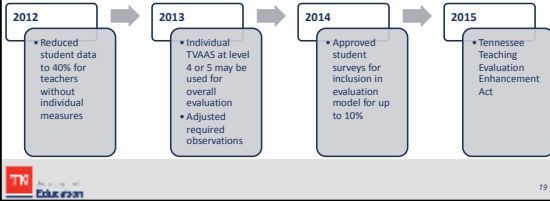
Teacher Evaluation

- **All school-based personnel, including administrators and support services**
- **Multiple measures including:**
 - Observation rubrics
 - Individual TVAAS or Portfolio Growth Measure
 - Other achievement data (AP, Graduation Rate, etc.)
 - Student survey (optional)
 - Teacher survey (evidence for administrator rubric)



Teacher Accountability

- Multiple measures evaluation system with five performance levels
 - Classroom observations (TEAM, TIGER, TEM, Project COACH)
 - School- or district-wide and/or individual growth data (TVAAS or portfolio)
 - Other student achievement measures (TCAP/EOC, TVAAS, graduation rate, AP results, etc.)
 - Student Survey data (optional)
- Annual report produced each year used to refine model



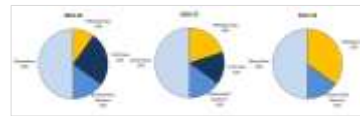
Changes During TNReady Transition

Tennessee Teaching Evaluation Enhancement Act

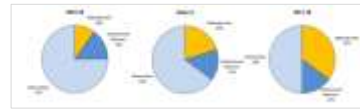
- Adjusts the current weighting of student growth data in a teacher's evaluation during the transition to new TNReady and social studies assessments.
- New state assessments will factor in a teacher's evaluation as follows:
 - 10 percent of the overall evaluation in the first year of administration (2015-16),
 - 20 percent of the overall evaluation in year two (2016-17), and
 - 35 percent of the overall evaluation in year three (2017-18)
- Student growth data for teachers in non-tested grades and subjects currently counts for 25 percent of the overall evaluation.
 - The weight will lower to 10 percent in 2015-16 and move to 15 percent in subsequent school years.

Tennessee Teaching Evaluation Enhancement Act

- Tested Teachers with Prior Data

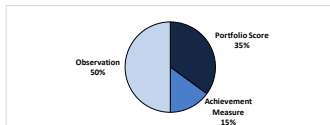


- Tested Teachers without Prior Data

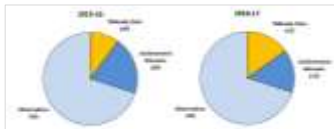


Tennessee Teaching Evaluation Enhancement Act

- Non-Tested Teachers Using a Portfolio Growth Model



- Non-Tested Teachers



Accountability: Closing Thoughts

Public Reporting

- School, District, and State level reporting via online **Report Card**
- Full **transparency** regarding:
 - District and school progress
 - Reward, Focus, Priority status
 - Achievement data by assessment, by subgroup performance
 - Participation rates
 - Graduation rates
- Individual teacher evaluation data (including TVAAS) protected from public release or open records requests

Accountability Alignment

Element	Teacher/ Administrator	School	District
Growth (TVAAS/Portfolio)	X	X	X
Achievement (Grad Rate, AP, other)	X	X	X
Gap Closure (Subgroup progress)		X	X
Student Surveys	X		
Teacher Surveys	X		
Parent Surveys			
Observations (Rubrics)	X		
Public Reporting		X	X

Progress in Tennessee

- Accountability has been one part of the systems change in Tennessee over the past several years.
- Accountability coupled with higher standards and aligned assessments will continue to move Tennessee students to achievement.
- “What gets measured, gets taught.”
- We must continue to consider ways to hold us all accountable for the outcomes that matter the most for our students.
- We should have a mindset of continuous improvement in all areas that impact student learning.



Questions?

For Taskforce Consideration

- What questions do you have about our accountability framework?
- Are there principles related to accountability that we should endorse as a state?
 - If so, what are they?

Assessment Practices Task Force: Meeting Five – August 24, 2015

Task Force Goals:

- To conduct an environmental scan of assessment usage and practices across the state.
- To establish principles addressing purposes and goals of state assessment relative to locally designed or locally chosen assessments (i.e. formative assessments).
- To define appropriate practices associated with these principles that best support decision-making at the state, district, school and teachers levels.
- To gain insight on ways to best communicate about TN Ready to all stakeholder groups.

Meeting Five Goals:

- Gather feedback on TNReady Reports
- Share information on work of the Scheduling and Logistics Task Force
- Review and gather final feedback on the task force report
- Conclude work of the task force

Topic	Components	Timing
Welcome	<ul style="list-style-type: none"> ▪ Review the agenda and goals 	8:30-9:00
TNReady Sample Reports	<ul style="list-style-type: none"> ▪ Review multiple versions of sample TNReady score reports and provide feedback ▪ Small group and whole group feedback 	9:00-9:45
Scheduling and Logistics Advisory Council Update	<ul style="list-style-type: none"> ▪ Update on work to date ▪ Discuss pros and cons of sample assessment schedules 	9:45-10:15
Break		10:15-10:30
Feedback on Report	<ul style="list-style-type: none"> ▪ Discuss feedback from task force members ▪ Share updated draft of report 	10:30-11:00
Report Communication Plan	<ul style="list-style-type: none"> ▪ Share Communications Plan ▪ Discuss next steps related to recommendations 	11:00-11:45
Closing	<ul style="list-style-type: none"> ▪ Wrap-up task force work ▪ Survey ▪ Final Comments 	11:45-12:30

Groups for Discussion:

Group 1	Group 2	Group 3
Candice McQueen	Nakia Towns	Vicki Kirk
Sara Heyburn	Dolores Gresham	Nicole Roberts
John Forgety	Wanda Shelton	Harry Brooks
Mike Winstead	Sharon McNary	Beth Unfried
Philip Eller	Becky McBride	Debbie Shedden
Jasmine Carlisle	Virginia Babb	Nancy Ashe
Danielle Mezera	Kathleen Airhart	Bill Harlin
		Valerie Love



Assessment Practices Taskforce

Final Task Force Meeting
August 24, 2015

Agenda

Time	Topics
8:30-9:00	Welcome
9:00-9:45	TNReady Sample Reports
9:45-10:15	Scheduling and Logistics Advisory Council Update
10:15-10:30	Break
10:30-11:00	Feedback on Report
11:00-11:45	Report Communication Plan
11:45-12:30	Closing

Reminder of Task Force Goals

- Conduct an environmental scan of assessment usage and practices across the state
- Establish principles addressing purposes and goals of state assessments relative to locally chosen/selected assessments
- Define appropriate practices that best support decision-making at the state, district, school, and teacher level
- Gain insight on ways to best communicate about TNReady to all stakeholder groups.

Review of Progress to Date

- Heard feedback on 10,000 Teacher Tour roundtables
- Reviewed survey results (SCORE and TDOE)
- Assembled task force
- Identified four primary Areas of Feedback
 - Summative assessment
 - Formative assessment
 - Test prep
 - Logistics and scheduling
- Reviewed presentations and data on areas of concerns
- Discussed possible solutions
- Drafted principles in four primary areas

Review of Progress to Date

- Drafted recommendations based on principles
- Worked on recommendations when possible
- Finalized report and identify other areas of analysis



TNReady Score Reports

TNReady will measure if students are **on the right track...**

- TNReady assessment will provide students, teachers, and parents with more accurate and authentic information about each student's progress and achievement.
- TNReady offers parents, students, and teachers a new and improved academic check-up each year to make sure all students are moving forward and are on track to graduate from high school and be successful in college and the workplace.



And provide better information for families

- Parents will receive **more and better** information on their student's performance.
- The new reports are designed to do three things:

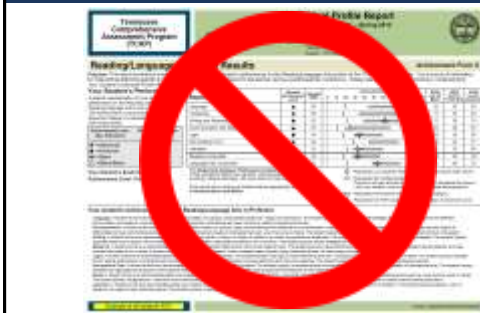


Reports will be designed to answer key questions

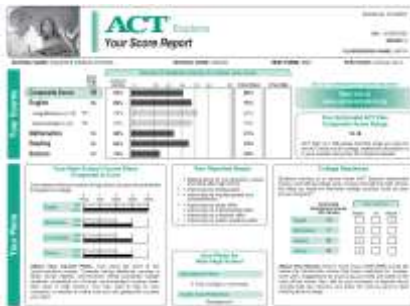
- *How is my student performing according to standards/expectations?*
- *What are areas of strength and opportunities for improvement?*
- *How well is my student performing relative to school, district, state?*
- *What can we (students, parents, educators) do now?*



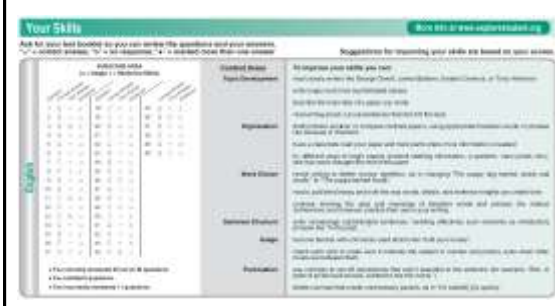
Current TCAP reports don't meet these needs



Many find ACT EXPLORE reports more useful

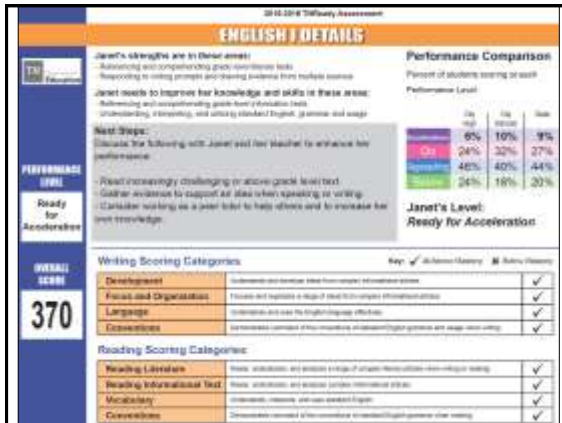
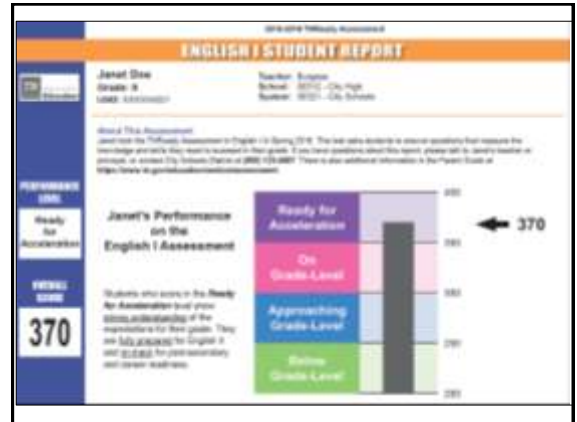


ACT EXPLORE reports provide information on **how to improve...**



TNReady report design is now in the beginning stages

- New report sample designs are targeted to reflect the feedback received through the taskforce thus far.
- In addition to the key questions and design priorities, we are considering content and layout elements:
 - Single subject versus portfolio (multiple subjects)
 - Performance level names (e.g. distinguished, ready for acceleration, advanced)
 - Performance level descriptors
 - Score scale ranges (below 500 versus above 1000)
 - Graphics (bars, stars, bubbles, lines, etc.)
 - Color scheme & orientation (portrait versus landscape)



Group Feedback (20 Minutes)

- Each group has two sample reports: single subject and portfolio
- Please discuss and critique the following elements:
 - Performance level names (e.g. distinguished, ready for acceleration, advanced, etc.)
 - Performance level descriptors
 - Score scale ranges (below 500 versus above 1000)
 - Graphics (bars, stars, bubbles, lines, etc.)
 - Color scheme & orientation (portrait versus landscape)
- Does the sample report adequately address the key questions:
 - How is my child performing according to standards/expectations?
 - What are areas of strength and opportunities for improvement?
 - How well is my child performing relative to school, district, state?
 - What can we do now?



Whole Group Discussion

- What were your overall impressions of the sample reports?
- What information is missing?
- Which did you like best about the samples?
- What did you not like about the samples?



Next steps: Better feedback for families

- **Fall 2015:** Parents, educators, and higher education faculty across the state to provide feedback on the format of the new reports, as well as performance level names that clearly reflect student progress toward success after graduation.
- **Winter-Spring 2016:** Educators and higher education faculty will create performance level descriptors that define what students know and are able to do at each level
- **Summer 2016:** Educators review student responses and set criteria to score at each performance level
- **Fall 2016:** TNReady reports delivered to educators and families





Test administration and scheduling is a real challenge

- Taskforce feedback has led to creation of scheduling and logistics advisory council
 - Composed of 18 districts and includes 6 elementary, middle, and high schools each
- Council established to provide feedback and tools related to test administration in the first year of TNReady transition
- Kick-off via webinar in late June, followed by in-person planning session on July 15
 - Additional webinars planned monthly, with in-person meetings scheduled for September regional sessions, LEAD conference in October, and January planning session

Flexible Administration

- Districts have **more flexibility** than ever before to choose when they want to administer the test
- Giving the test online gives districts more flexibility to schedule the test around instruction as opposed to *shutting down* the school for testing
 - Students will *not* be testing during the entire testing window
 - All schools in a district do not have to test on the same day
- Exemplar schedules identified and released to provide models of testing schedules that minimize disruption
 - Include a K-5, K-8, 6-8, and 9-12 school
 - Schedules demonstrate a variety of approaches
 - School characteristics providing for context comparisons
- Additional sample schedules provided at each grade level beyond exemplars

Exemplar Schedule – K-8 School (Putnam County)

The image shows a detailed calendar-style schedule for a K-8 school in Putnam County. The subject is '4th Math & 3rd ELA'. The schedule is for Tuesday, February 9th. It shows a grid of time slots from 7:00 AM to 3:30 PM. Specific blocks are highlighted in blue and orange, representing different testing periods. To the right of the grid, there are notes and instructions regarding the schedule, such as 'Please do not change the regular schedule' and 'We will not change the regular schedule'.

Exemplar Schedule – 9-12 School (Blount County)

The image shows a detailed calendar-style schedule for a 9-12 school in Blount County. The subject is '9-12th Grade Geometry & Algebra'. The schedule is for Tuesday, November 15th. It shows a grid of time slots from 7:00 AM to 3:30 PM. Specific blocks are highlighted in yellow, green, blue, and purple, representing different testing periods. To the right of the grid, there are notes and instructions regarding the schedule, such as 'Blount County/Heritage High School' and 'We are using accountability to justify the order we are testing subjects.'


Scheduling Toolkit

- **2015-16 Assessment Calendar template** – a calendar that gives an overview of all of the testing windows for strategic planning
- **TNReady Testing Schedule template** – a generic template that can be used to plan each school day during an assessment window, taking into account times, technology, and locations available
- **TNReady Scheduling District Decision Tree** – a series of questions to help district level administrators make informed decisions about assessment scheduling
- **TNReady Scheduling School Decision Tree** – a series of questions to help school level administrators make informed decisions about assessment scheduling
- **Calendar Cover Sheet** – this document can assist in gathering the information needed for scheduling decisions
- **Business Rules / Best Practices for Testing** – this document summarizes the business rules of scheduling, as well as provides some "best practices" for testing that affect scheduling decisions




Recommendation 1

- TDOE should continue to focus on **improving communication around testing and accountability** to create clarity, transparency, and trust.
 - Next Steps:
 - Communicate TNReady to media and general public audiences
 - Collaborate on communication, training, etc. with all education stakeholders and partners
 - Create proactive and clear communications to educators
 - Post additional information about TVAAS on TDOE website




Recommendation 2

- Beginning with the 16-17 school year, the TDOE should **annually release as many summative test items as possible** without compromising test security and development. These should include operational, non-linking test items.
 - Next Steps:
 - Determine cost
 - Create plan for public release




Recommendation 3

- TDOE should **annually release standardized test blueprints, test specifications, and the methodology** for calculating reports.
 - Next Steps:
 - Have already released these items for this year
 - Prepare clear timelines for release in future years
 - Staff appropriately




Recommendation 4

- TDOE should convene a **testing scheduling and logistics advisory group** that is representative of district personnel across the state to address logistical challenges. This team should focus on test scheduling that minimizes disruptions and provides ongoing guidance and support (in the form of sample schedules and process guides) particularly during the first year of TNReady.
 - Next Steps:
 - Already convened group
 - Shared initial samples with assessment task force
 - Will share with districts and provide ongoing technical assistance
 - Analyze results for continued improvement



Recommendation 5

- TDOE should ensure annual tests **provide clear reports** for educators, parents, and other stakeholders that point to alignment to postsecondary readiness.
 - Next Steps:
 - Created samples for feedback
 - Shared with assessment task force
 - Will share with additional groups



Recommendation 6

- Along with other groups working on the test reports, a **parent advisory group** should be formed to give input and feedback on report options.
 - Next Steps:
 - Started collecting potential names for parent advisory group



Recommendation 7

- TDOE and individual districts should build upon current reporting requirements related to mandated assessments and **clearly communicate to the public the purpose of large-scale formative or summative assessment usage**. If either TDOE or an individual district administers a large-scale assessment, teachers, parents, students, and other stakeholders need to know "why," with such information easily accessible to the public through district and school websites, as well as other sources.
 - Next Steps:
 - Posted information on state annual tests on TDOE's website (also created educator and parent guidebooks on TNReady)
 - Will present principles/this recommendation of Assessment Task Force to directors at conference during September
 - Will determine method to ensure compliance



Recommendation 8

- TDOE should **create additional portfolio options** for teachers in non-tested grades and subjects, specifically for first grade, from which districts can choose.
 - Next Steps:
 - Already started first grade portfolio process that will result in pilot in spring 2016
 - Examine potential for additional options



Recommendation 9

- TDOE should create its **own second grade assessment aligned to Tennessee state standards** as an alternative to SAT-10 and require district administration in an effort to improve statewide literacy.
 - Next Steps:
 - Drafted RFP to begin process



Recommendation 10

- TDOE should continue to require ACT for all 11th grade students except for the rare circumstances in which an IEP precludes a student from taking the ACT; however, to address the issue of over-testing, the state should remove the mandatory EXPLORE (8th grade) and PLAN (10th grade) tests, and not adopt ACT's new alternative ASPIRE, IF reports from TNReady or other methods can point to similar information.
 - Next Steps:
 - Continue to ensure standards alignment to ACT and create TNReady based on postsecondary readiness standards
 - Explore options to provide students and parents with career inventory information



Recommendation 11

- The department should consider re-purposing funds currently expended on the discontinued 8th and 10th grade EXPLORE and PLAN tests for **one ACT retake** opportunity for interested students.
 - Next Steps:
 - Explore options in budget
 - Explore how to regulate at the state level



Recommendation 12

- TDOE must ensure future and new teachers have explicit knowledge about assessment for learning. Tennessee teacher preparation programs should include a **specific curriculum or module on assessment with specifics to Tennessee**. All new teachers also need specific training with embedded professional learning around this area of the TEAM rubric.
- Next Steps:
- Create advisory group to determine contents
 - Determine policy and process changes



Recommendation 13

- TDOE should communicate with educators, through regular channels including regional assessment meetings, about **best practices for test preparation for annual tests and those that should be avoided**. Communication should center on the idea that the best test preparation is focusing on great teaching and engaged student learning every day.
- Next Steps:
- Scheduled regional assessment meetings for September
 - Create messages, guidance documents and communication for districts and educators about principles of assessment



Recommendation 14

- TDOE should **assist districts in transitioning from using school counselors as test coordinators** so that they are available to deal with the needs of children during periods of testing.
- Next Steps:
- Communicate expectations of assessment principles to directors at Superintendent's Conference in September



Recommendation 15

- TDOE should work directly with districts to **increase awareness of the realities of test anxiety** while providing specific guidance in how to help educators avoid passing on stress or test anxiety to students. School counselors should assist in this work.
- Next Steps:
- Communicate expectations of assessment principles to directors at Superintendent's Study Council Conference in September



Communication Plan

Introduce Task Force Members

- Feature members of the group on social media
 - Include picture, biographical information, and a quote about the work of the task force
- Highlight members on both Twitter and Facebook
- Share with legislators to also promote through their social media channels



Introducing **Becky McBride**, a member of the Tennessee Task Force on Assessment. Becky is the 2015 Southwest Region Teacher of the Year and was a state finalist for Tennessee Teacher of the Year. She teaches English at Brighton High School and holds multiple teacher leader positions.

Becky says working with a variety of stakeholders has energized her about the future. "As the fastest improving state in the country, Tennessee has accomplished so very much in the last ten to fifteen years, and I am humbled at the opportunity to be a voice for teachers across our state in regard to assessment and its many realms."



Timeline for Release

- Tentatively planning to release the report on Sept. 3
- Coordinates with the kick off of the Commissioner's *Classroom Chronicles* Teacher Tour
- Share with members of the General Assembly
- Promote on social media and our blog, *Classroom Chronicles*
- Present findings to the Superintendent's Study Council Conference



Post on our Website

- Highlights, as well as the full report, will be posted on the department's assessment webpage:
<http://www.tn.gov/education/section/assessment>



Questions?

51

Discussion Groups

Group 1	Group 2	Group 3
Candice McQueen	Nakia Towns	Vicki Kirk
Sara Heyburn	Dolores Gresham	Nicole Roberts
John Forgety	Wanda Shelton	Harry Brooks
Mike Winstead	Sharon McNary	Beth Unfried
Philip Eller	Rebecca McBride	Debbie Shedden
Jasmine Carlisle	Virginia Babb	Nancy Ashe
Danielle Mezera	Kathleen Airhart	Bill Harlin
		Valerie Love



Appendix 2

Minutes

Assessment Task Force
Meeting One - April 6, 2015

- Welcome from Commissioner McQueen and introductions

- **Goals:**
 - Discover what tests are being given across the State
 - What are goals of state assessments? Local assessments?
 - Define appropriate practices
 - Ways to best communicate about TN Ready
 - While on the 10,000 Teacher Tour, over and over teachers were concerned about the number of tests and were confused why they were all necessary. This taskforce is a response to feedback.
 - Multiple states looking at this issue

- **Final Product:**
 - Compiling a report with recommendations that could be revisited over time

- **Additional Goals?**
 - Hard look at state assessments and local assessments going to take place
 - The logistics of testing:
 - How many days do the tests take? (Time, duration)
 - The context of the school, the size of the school can effect
 - Weather can interfere with instruction making it more complicated
 - Professional Learning Communities – how does this fit into supporting decision making fit in? Need for PLCs?
 - At the school level: using data – how? Why?
 - There is not good clarity to parents about testing – the score is given, but parents don't understand what it means, what it is for, and how is used to help their child
 - Special education students – given the same type of state test as a general student at end of the year even though we have given accommodations to them all year

- **Vision for the end product?**
 - Students who take ownership of learning – understand why and what, can compete on an international level
 - Teachers feel less stressed about testing – not feeling so overwhelmed; learn from other teachers in different districts: what worked? What doesn't? Create a connected community of teachers – teacher ownership
 - Responsive to federal issues

- We have to create buy-in by communicating with all types of people (parents, legislators, principals, etc.), so when we get results, people don't discount them and we are back at step one
 - Before it is final, meet with legislators and give them an opportunity to express their concerns on the front end – raise awareness among legislators
 - There are so many misperceptions, and “perception is reality.” People think that most assessments are state-issued, and they raise their concerns to legislators. Those legislators then create legislation around those concerns. We need to make people know exactly what the State asks for so people are aware that everything else is a LEA decision.
 - Principles for how LEAs need to make decisions around testing need to be clarified (district level clarity)
 - Get back to making tests for the student meaningful
 - Principles/guidelines for districts to make assessment decisions but **not** mandates
 - Better implementation for assessment training
 - Look at the technology piece as well (3rd graders don't have keyboard skills)
- **Today's Goals:**
 - Review state landscape of assessment in Tennessee
 - Review district survey
 - Gain feedback on district survey and teacher/student surveys
 - Knox County is also doing a study on assessment, but we are aware and there is not any overlap. (They are the only county doing so.)
 - **Emily Freitag and Nakia Towns – Presenters**
 - The purpose of assessments are varied:
 - Some compare performance, some place students on a spectrum of performance, some give a baseline, some measure a goal, some are used to plan instruction, etc.
 - **One of themes of this task force will be purposefulness**
 - Is a test valid and reliable?
 - Validity – how well does the test accomplish the purpose?
 - Reliability – is this test telling us what students know? Stable? Consistent?
 - Skills-based: documents students' performance of a specific skill
 - Standards-based: documents how well students know the standards
 - Work with experts and statisticians to ensure this

- **National Assessments**

- ACT/SAT – TN typically gives ACT, but they have the option
 - Source: It is in state law
 - Purpose: Measure college readiness, determines HOPE eligibility
 - Cost: \$45.25 per student
 - Not directly reimbursed: Funded through BEP funding (Based on # of graduating students that the school had the previous year)
- EXPLORE (All 8th graders) / PLAN (All 10th graders)
 - Source: In state law
 - Purpose: To gauge college readiness
 - ACT is phasing this out. This will be given next year, but then it is up for debate. This would be a good thing for this group to discuss
 - “Does this test go beyond college?” – Some post-secondary is needed for most careers, so post-secondary readiness (in a lot of ways) is career readiness
 - What does career readiness mean? How do we define it?
 - Being discussed in the DOE’s strategic plan
 - Could TNReady fulfill this requirement? We are going to have to examine this more because these tests are going away
 - Cost: \$10.50 per student (Comes through state budget)
- NAEP
 - Source: Federal requirement
 - Purpose: Common metric across states to measure relative performance and monitor education progress over time
 - Administration:
 - Not given to every student, given every two years
 - Different accommodations for NAEP for special needs students
 - Periodically administered in specific subjects: arts, civics, writing
 - Don’t get district level results – just state results
 - Cost: does not cost the TN any money

- **State Assessments:**

- History:
 - TCAP began in 1983 (Mandated in 1992)
 - TVASS also began in 1992
 - Timeline can be found in PowerPoint slides

- TCAP/EOCs (Untimed) :
 - Source: Federal and state law
 - Writing portion is now required by state law for grades 3-11
 - Purpose: Measure mastery of grade level (or course) standards
 - Cost: \$20-25 (No cost to districts)
- TCAP changes are in 2015-2016
 - Physical Science – there is no plan for a statewide assessment for this
 - Math, RLA, and Writing TCAP tests will be replaced by TNReady
 - Math: 3 subtests, one administered in February, two in April
 - ELA: 4 subtests, two administered in February, two in April
 - There is a high school block of administration and secondary block of administration – flexibility
 - Before TCAP, teachers stop and review; will this happen twice now?
 - Hopefully not because TNReady administration is less interfering than TCAP to the normal day. State is communicating that the best way to prepare for TNReady is to continue normal instruction
 - Blueprints already been published – teachers know how to pace their lessons to prepare their students for the test
 - Opportunity to rethink what test preparation really means
 - Test preparation (testing culture) is taking up more time than the actual test
 - Trying to keep each test 60-90 minutes so it can fit into a normal schedule so schools don't have to create a specific testing schedule
 - School flexibility on when tests are being taken: This is broken down to where each school can make the decision, as long as it is in the testing window
 - There is going to be an element of reorganizing your environment where it is not normal when we test – stringing this out may worsen this
 - Want flexibility but also want be able to eliminate logistical challenges
 - Think about overcrowded schools – teachers on carts – having to change around instruction rooms during testing
 - Online testing is going to eliminate a lot of logistical problems, as is the type of test TNReady is going to be
 - No shipping and securing paper in buildings

- No more butcher paper on walls
 - This test will not allow any information on a poster will allow you to answer a questions
- TCAP – Alt (Portfolio will be replaced by NCSC)
 - More expensive than normal assessments (\$127.61 per student)
 - This follows closely with other states
 - Administered to 1% of students excluded from general state assessment
 - Currently all four subjects – NCSC will replace ELA and math, will maintain for social studies and science
- WIDA/ACCESS for English Learners
 - Source: Required by federal law
 - Purpose: Evaluate English proficiency
 - Begins when a student starts school, continues until the student reaches a level of proficiency that deems them no longer an English learner
 - Cost: \$25.75 per student
- K-2
 - Optional assessment (with a declining trend in participation)
 - Alignment is the largest voiced problem with this test
 - SAT10 is being phased out after 2015-2016
 - Teachers have expressed that they would rather have a criterion-reference test (rather than the current norm referenced format)
 - Potential RFP for 2nd grade option
 - Is this actually a choice? Because if it isn't taken, then those teachers' growth scores come from their feeder schools.
 - It is a state requirement that 3rd graders receive intervention if they are behind, which is why it is important to have good data
 - In the waiver application, the State proposed students taking Algebra 1 in the 8th grade would count towards the middle school accountability and high school accountability in scores (Scores counted twice)
 - We will not know until later if this is approved
 - “We need benchmarks for kindergarten and first grade to see where those students are.”
 - Cost: \$20.10 per student
- RTI:
 - Schools can choose which screener, but a screener must be used

- More information can be found in PowerPoint slides
 - Often districts give assessments for “predicative validity” purposes which is a huge problem going forward. This traditional method will not be useful because, without a few years of TNReady data, we don’t know what to test
- Alyssa Van Camp (SCORE) – Presenter
 - Engaging Peabody College of Education to help analyze data in survey
 - Big questions:
 - What is currently being implemented?
 - How is the data used?
 - How much time is spent testing?
 - Going to send the survey in a link with targeted questions to get feedback
 - Going to send this out this week and will want it back by the first week of May
 - This is going to district leader only – a different survey will be sent to teachers and students later
- **Feedback:**
 - “How much time does it take to administer?” Hard to calculate because it depends
 - This needs to be clarified – total time or student test taking time, etc.
 - What is your block of time to give it? How long does it actually take to finish?
 - We need to clarify that this is an overview as opposed of hours of finding data. “What do you already know?”
 - Several tests can be given as paper/pencil AND online
 - Clarify which are state assessments (Put an exact list)
- **Timeline:**
 - April: Administer all surveys
 - May: District interviews and focus groups
 - June: Data analysis and report writing
 - July: Report review and release
- **Future Meetings:** (Exact times are coming)
 - May 20 – Full Day
 - June 16 – Morning
 - July 15 – Morning
- **Homework:**

- Read article provided in packet
 - Choose stakeholder group (parents, students, teachers, etc.) and choose three people in this group – get their perspective on the questions provided in packet
 - Will be provided with other states' information who have studied this
- **Next Steps for May 20 meeting:**
 - Review results of surveys
 - Review usages of assessment results
 - Design principles for purposes and goals of assessments
 - Begin drafting recommendations

Assessment Task Force Meeting #2

May 20, 2015

- Commissioner McQueen welcomed group, reminded group of goals
- Emily Freitag, Assistant Commissioner of Curriculum and Instruction
 - Broke large group into smaller groups to have discussion about the homework
 - What did you notice between the difference of interim, formative, and summative assessments?
 - Some educators don't realize that several of their classroom tests are actually summative (because you give the test, give a score, and move on) instead of an interim test
 - We need to teach teachers more about the different types of assessments– (This is a professional development opportunity); “these are tools, not to catch people”
 - This is a culture shift for parents and teachers
 - When we just throw numbers and proficiency levels at students, without explaining the purpose, it does not mean as much as it should
 - “Interim was missing from my vocabulary;” it helped me better understand what exactly I was already doing in my classroom
 - It is very cumbersome for teachers of older students (who are being tested on more) to drill down the information to find which students know what – need better tools to do this
 - “Students think assessments are pointless because they aren't getting anything out of them except a grade.”
 - Emily read the definitions of interim, summative, and formative assessments found in the article so everyone is using the same vocabulary
 - Broke large group into smaller groups to have discussion about their interviews
 - What did you hear? Trends?
 - Misunderstanding of scoring, and populations that are scored
 - Students don't know the purpose of the assessments, aren't receiving real feedback on benchmark, classroom tests
 - Most students think you should shorten/give more time – this would make assessments better
 - Teachers think there is too much testing
 - Central office staff/superintendents are excited about TNReady but districts did not want their hands tied about assessment decisions
 - 70% of test questions should be released since they have to be refreshed anyway; could put them out on the website
 - Misunderstanding – parents don't read state reports because they don't understand what is happening

- Miscommunication – Acronyms are difficult, even for teachers; teachers can't communicate about previous year's test scores because they are already with a new teacher
 - Education speak vs. education policy speak vs. normal people speak
 - Mistrust – when parents can't see what their children missed, they don't trust it; more trust around teacher driven testing
- Students take an exam the week after they take an EOC, is this contributing to too much testing?
 - Honors students must take EOCs (which are regular level) then take their finals
- Stress that trickles down from teachers to students is a problem
 - This is especially bad for teachers that have very high level students that are difficult to grow
 - The level 4 and level 5 teachers are hesitant to take student teachers because the tests are so high stakes – “These are the best teachers, and they should take be taking student teachers.”
 - How do we manage the stress and anxiety that teachers are feeling? How do we create clarity about what assessments mean?
 - “Kids feed off your stress level.” – Some teachers are treating assessments as “celebrations of learning”
 - Not every teacher feels stress about teaching – we don't need to think only in those terms in this task force.
- Alyssa Van Camp, SCORE Policy Director
 - For full presentation, see the folder you were provided in the meeting:
 - District Survey – 48.9% response rate
 - Purpose (top 3)
 - Tracking student progress
 - Diagnosing student skill deficient
 - Inform teacher support practices
 - Challenges (top 3)
 - Scheduling
 - Student prep for technology assessments
 - Technology availability
 - All of these are logistical issues; we are hearing this issue over and over again in focus groups as well

- Principal Survey – 11.1% response rate
 - Purpose (top 3)
 - Diagnosing student skill deficient
 - Predicting student performance on future assessments
 - Inform teacher support practices
 - Challenges (top 3)
 - Scheduling
 - Technology availability
 - Selecting high-quality assessments for my school
 - Logistical issues also are being felt at the school level
 - Principal responses were higher than district leaders as they reported the number of assessments being given, and the amount of time spent (Slightly higher percentage of them thought that there were too many/too much)
 - Do districts and principals think they can do anything about the amount of testing?
 - “I think there is too much, but I don’t feel like I can get rid of anything.”
 - “If it wasn’t required, would you go back to not having the assessments?”
 - “I wouldn’t test my students in the 95thile three times a year.”
 - Do we need district wide aggregate data?
 - Can we find an assessment that serves multiple purposes? Would this help with the balance?
 - Are principals not giving school level assessments because of the number of mandated assessments?
 - This could be a bi-product that
 - District driven, purpose oriented assessment is very important
 - Teachers are asking for assessments, possibly because it is difficult to make assessments – Is there too much trust in vendor tests?
 - We need to empower teachers to create assessments because there is more buy-in and understanding
 - “I think districts want someone to check their work.”
 - Leadership is key in how assessments are interpreted
- Teacher Survey – 14.9% response rate
 - Purpose (top 3)
 - Diagnosing student skill deficient
 - Informing instruction
 - Set goals with students

- Challenges (top 3)
 - Reduced instructional time
 - Technology availability
 - Student prep for technology assessments
- Compared to districts/principals, teachers significantly believe that there are too many assessments and that they spend too much time of them
- Amount of time spent preparing for district assessments falls on the opposite ends of the spectrum
 - Focus groups indicate that this is highly dependent on the way the leadership thought about testing
- What did the responses look like with different grade levels?
 - Has not been cross-tabbed yet, but that can be done
 - K-8 are tested more
 - Remember that this is not a fully representative sample, so we need to take the data seriously but look at it with a careful eye
- Takeaways – slide in PowerPoint
- Alyssa broke large group into small groups to talk about survey findings
- Share out from small group discussion:

- **Teachers:**

- Challenges:

- Too much assessment
 - There is a benefit, but there is too much
 - It feels separate from learning
- Measurement on evaluation
- Logistics vs. Content
 - Technology
 - Time
- Secondary – Whole school shuts done when one test is being given
 - “We have to stop learning for everyone.”
 - One day in May when there was no assessment
- Physical classroom environment changes
 - TNReady rule changes will help change this
- Use of data

- Solutions:

- Implementation of TNReady
- Give students opportunities to engage in technology
- Consistency for transient students
- Needs to be more intentional support for teachers
 - Professional development on using data to effectively make instructional choices

- RTI
- Assessments should be student-focused, transparent for all stakeholders
 - Full disclosure, over-communicate the clear purpose
- Norms/principles:
 - Logistics – minimize challenges at district level
 - Collaborative – Conversations around purpose of test
 - Intentional support resources
 - Student-focused feedback
 - Focus on growth of all students
 - Transparency for all stakeholders
 - Use assessments to grow students
- **Principals:**
- Challenges:
 - Scheduling
 - Scheduling assessments is a big concern – giving the actual assessment; the logistics in a building; unfortunate; assessments are not conducive to class periods; get folks off that normal pattern
 - Also students in special education – have some pull out things
 - The computer based nature makes it challenging, but we have been working on the technology because we have ramped up the technological aspects; just to protect instructional time;
 - Platforms were a challenge in particular for social studies
 - Technology
 - Need opportunities for students to engage in technology
 - It is not only the equipment – it is the exposure; with keyboarding and dropboxes – need a lot of experiences – to make sure assessing the content; that is another facet and the platform
 - We need to fund technology for the purposes of day to day learning; most of us are three to one
 - BEP challenges
 - Variability from district to district – lower socioeconomic
- Top uses:
 - To predict student performance disparity – for principals are highly accountable; any manager wants to know where am I tracking
 - Challenges with vendors – people might be flying blind for a year
 - Predicting skill deficits – have an idea if they are within the district; but for transient children it is important

- I look for teacher variability; you look at student performance but also looked for trends for teacher effectiveness – principal perspective – looking for that pattern
 - Think that teacher variability might be a focus for school leaders
 - This is a tool for you – to inform your instruction – to get you support; they see the value versus
 - For the longest time we just had the autopsy effect – I think teachers like taking ownership
 - It will be interesting to see how RTI impacts this – skill deficits; probably elementary teachers responses
- Solutions:
 - Need to find technology and providing opportunities for people to practice
 - Technology comes into play in every discussion
 - Technology doesn't last forever
 - One solution could be supporting tech directors and a BYOD strategy
 - The issue is the keyboarding, platform, and size of the screen
 - Variability – security of the testing system
 - Think it will be a different situations
 - When put out those RFPs or bids
 - How much are our standards going
 - Easy wins in scheduling support, coaching principals, keep the test length manageable might help (won't totally fix the problem)
 - Practices and Norms:
 - Discussion around formative and summative; what is our belief versus a district belief; driving it in a principal; common conversation
 - Communication around the purposefulness of testing
 - Framing the big picture – all in how the leader presents the assessment strategy; positive and intentional – principal is able communicate – principals and central office leaders need some support
 - Parent belief?
 - Seems to be a disconnect
 - Action plans for principals around balanced assessments
 - High performing schools that have had movement
 - Clarity to principals; challenges with the accountability; we have huge gains; when we are clear with what we are going to be assessed on and held accountable;
 - Release items with TNReady is a positive and a good example
 - Real opportunity to set the conversation with the rollout with TNReady

- **Districts:**
- Challenges:
 - Teacher buy-in
 - Once you have teachers bought into the assessments, everything takes care of itself.
 - Dr. Winstead: Teachers like assessments in our district because they do well, it's another opportunity for them to celebrate that. It's a different experience in districts that aren't as high performing and ends up being one more piece of data that highlights weaknesses.
 - Accountability – Everything about assessments is colored by this
 - Focus on growth – competition emerges but we need to compare to ourselves (Are we getting better?)
 - It's about measuring growth over time. It's not about comparing yourself to other students, teachers, schools, etc.
- Solutions:
 - Transparency:
 - Districts and TDOE should be fully transparent with all stakeholders about TNReady—explain every aspect of the assessment and give them a good idea of what's going to happen each step of the way.
 - Ensuring that communications gets to the school and teacher level—it often breaks down before that.
 - Over-communicating about assessments—make sure to hit it in summer PD, send lots of emails, etc.
 - Always communicate the purpose of assessment and explain the results.
 - Focus on growth
 - Always focus on growth measures with formative, interim, and summative assessments.
 - Eliminate redundancies (Clear purpose)
 - Annually conduct an assessment inventory at the district level to determine the purpose of assessments and eliminate any redundancies.
 - Norms/principles:
 - Focus on growth, not proficiency
 - Transparency – clear understanding of purpose, explanation of results
 - Addressing feeling of feeling over-tested
 - Annual assessment inventory could be an option to increase the frequency of the conversation

- Commissioner McQueen presents 2007 freshmen cohort information, strategic goals
 - 55% of TN students are on free/reduced lunch (from a disadvantaged background)
 - 63% of students take a remedial course in the first year of college
 - 48% of students drop out in their first year of college
- Nate Schwartz, Director of Research and Policy - Full presentation can be found in folder
 - What is a realistic goal to aspire to on the %age of Tennesseans earning a 21 on the ACT?
 - We don't know, but the scoring of the ACT does not prevent us from getting extremely high percentages
 - Among the states where all students take the ACT, Tennessee is 4th out of the 12 states (Utah is #1 with a 20.7 average → moving to a 21 would take TN top the front of the pack)
 - The math proficiency progression table hides the story that we have way more advanced students in math, but the reading data does not hide any stories. It is flat.
 - Why are we moving off MAAS?
 - We are trying to build in accommodations into the new test for all students
 - Federal guideline: There is 1% latitude for these tests (99% of students must take the same test) - The 1% is assessed by portfolio.
- Emily Freitag, Assistant Commissioner of Curriculum and Instruction
 - TNReady (tee-n ready) Presentation – for full presentation, see folder
 - TCAP is not a good indicator of readiness
 - Could be doing well on TCAP and have to do remedial work
 - UT and TBR will be using TNReady results to place students
 - TCAP is the umbrella –
 - Achievement Tests : 3-8 grade
 - EOCs : 9-12 grade
 - TNReady will replace both of these in ELA and math
 - Math priorities:
 - Basic math skills are missing in our workforce
 - “Math people are made, not born.”
 - ELA Priorities:
 - 4 year gap between senior year in college and freshman year in college in terms of complexity of text
 - Stair casing complexity to prepare students appropriately
 - One of the biggest changes from TCAP:
 - Multiple different types of questions (not just multiple choice)
 - TNReady Tools:
 - TNReady Item Sampler – MICA
 - Much like item sampler of the past
 - Will be accessible from any computer with internet access
 - TNReady Practice Test – MIST
 - True simulation of full TNReady test

- Similar to practice test offered for writing assessment
 - These will both be available to all districts at no cost
 - Emily demonstrated MICA
 - Who makes the questions?
 - Measurement Incorporated develops first set of questions:
 - Three rounds of Tennessee Teachers review:
 - Round 1: Content
 - Round 2: Special Education Concerns
 - Round 3: Sensitivity
 - The questions then go to Forms Review
 - The MIST platform is “fixed form,” MICA is more flexible
 - Thoughts on TNReady:
 - Definitely more rigorous, but that’s what we want
 - Test prep is going to be a challenge; communication is key
 - Attack from the angle of the standard, not the tool
- The Social Studies field test was rare, but we had to do it because we didn’t have a year of testing
- You can find blueprints of TNReady
 - This is an effort to be more transparent
- Nakia Towns, Assistant Commissioner, Data and Research
 - TNReady will be administered online
 - Having 1 computer to 7 students typically means you are prepared for electronic testing
 - It is more expensive for paper-based test administration
 - MIST will hopefully be the online platform for all assessments
 - Developing technology literacy is very important
 - The department has committed to making the test fit into a normal class block
 - 45-60 minutes
 - EOCs will be timed (This has not been the case in the past.)
 - Quick scores were meant for grades only. They are not going to reflect proficiency or to be used for accountability. (They have in the past, but this is not going to happen moving forward.)
 - We are thinking what we will do with quick scores in the future
 - TVASS compares students to themselves, so even though their proficiency may change, the teacher score should not be effected when TNReady is implemented
 - Gathered information from Achieve Inc. to help communication plans
 - See full list in presentation
 - We had the best intentions when we designed the TCAP scorecard, but our reporting is unclear
 - Completely redesigning score card
 - Descriptions for what will be reported is still in developmental stages
 - Discussion of “Advanced, Proficient, Basic, Below Basic” verbage

- Small group break out to discuss the principles that should be put forth. (Principles will drive the recommendations of the task force.) Below are the report out notes:

Commissioner McQueen, Nathan James, Chairman Gresham, Director Winstead, Nancy Ash, Becky McBride, Barbara Gray, Danielle Mezera

- Standardized Testing:
 - State testing should show proper alignment to state standards.
 - Annual state tests are important. Student growth is necessary and a statewide annual assessment informs progress.
 - Grade-level annual assessments show progress toward postsecondary goals.
 - State assessments should help bridge expectation and communication gaps among teachers, schools, districts and the state.
 - Annual testing can better prepare our students for postsecondary entry.
 - Annual tests can help us continually improve to help us know if we have a world-class education.
 - Annual assessments points to level of rigor needed in the process of teaching and learning. Can define adult responsibility.
- Formative Assessment
 - Formative assessments that are teacher-led best inform instructional changes and these are preferred to general, large-scale possibly unaligned off-the-shelf assessments.
 - The number of formative assessment items should match goals. Too many test items are not encouraged. Formative assessments should have clarity and purpose. Formative assessments should be used for specific decision-making and next steps. Each formative assessment should have a clearly communicated purpose that all stakeholders understand.
 - Formative assessments should be used to measure mastery on standards that have taught. This does not preclude giving pre-assessments
 - Formative assessments should be designed in a variety of forms, not just paper-and-pencil (look at academic and soft skills). Intentional cross-curricular and multi-purpose formative assessments are encouraged.
- Test Preparation
 - Testing strategies should be part of the regular teaching and learning cycle. We should have unrelenting high quality instruction.
 - Technology must be integrated in the teaching and learning cycle on a regular basis. This is preparing students for the digital age.
 - The state, districts and schools must work toward daily technology usage for every student.
 - Absolute understanding of standards

- Test Logistics
 - The test should provide maximum flexibility in all testing decisions.
 - The state should provide more test support with opportunities for input from those on receiving end of tests.
 - Testing decisions should find proper balance between testing time and instruction.

Emily Freitag, Stephen Smith, Chairman Brooks, Virginia Babbs, Beth Unfried, Val Love, Bill O'Donnel, Vicki Kirk

- Standardized / Summative / Annual:
 - It helped me to know how we were doing with every student and predict their readiness for the next level.
 - See progress and growth year to year.
 - Show their culminating understanding.
 - Same measuring stick with other states or similar states other than just NAEP. This is important – gives us something we have in common. The public is crying for this because they will be competing.
 - Should provide information to students about strengths and areas in need.
 - You have to have something that gives everyone a good picture of where they stand.
 - Should provide precious information to the teacher and the future teacher.
 - Results be provided in a quick manner
 - The community needs to know as much about the test as possible and the state should release questions every year to create clarity, transparency, and trust.
 - We need to be able to learn from others to improve the curriculum.
 - It is required by law – federal and state.
 - 3rd grade requires it before you can move to fourth grade.
 - All students should take it otherwise you eliminate accountability and remedial planning – would destroy all these programs. Can't get a good handle on gaps.
 - Most important thing the state does – it's economics – we have to invest in making sure this is a priority
 - Scores should be connected to future opportunities
 - Assessment is part of the instructional process and the cycle of learning. Not a punishment.
 - Reports should be clear and provide timely feedback.
 - We talk too much about the test as the outcome and have to remain focused on the bigger goal of preparing students lives.
 - Reports should be understandable.
 - Tests should be short enough to fit in a normal class schedule but students should not be rushed.
- Formative Assessment / Interim / Benchmark / District:
 - Students deserve feedback about their progress along the way.
 - Every assessment should have a purpose that is meaningful for students (assessments should not be given just for teacher evaluation.)

- Results should be shared with students in an immediate way. Students should understand the purpose and next steps and should be connected to building skills for their future.
 - Districts should have flexibility to design their own programs but should also have support figuring out what the best plan looks like.
 - Districts should make sure their tests actually measure what the test is designed to measure.
 - Encourage collaboration in looking at the results.
 - A review of grading practices should be conducted.
 - If you are going to do formatives they should have these basic tenants.
 - Alignment to standards
 - Need for what assessment can look like – norming language, need for assessment and teacher prep and how to write great question
 - Teachers should ideally be involved in design, analysis, and review
- Test Prep:
 - Teaching the standards should be the best preparation – the best test preparation is an absolute understanding of the standards and how they will be assessed.
 - Do not spend the week before reviewing – children are bored to tears
 - Teachers should take care not to pass on stress
 - Students should have the opportunity to develop familiarity with the test platform in a natural way and in instruction
 - Task predicts performance
 - Students should not study test taking skills in isolation but embedded in the learning throughout the year
 - Technology should be embedded in instruction on a regular basis
- K-2:
 - It is way too late to know in third grade. Teachers need to know.
 - Best teachers must be in early grades
 - It has all changed
 - What my child is doing as a first grader
 - K, 1 and 2 is where it is at!
 - Interpretation of the standards is not quite there

Nakia Towns, Alyssa van Camp, Chairman Forgety, Debbie Sheriden, Bill Harlan, Philip Eller, Jasmine Carlisle, Kathleen Airhart

- Summative Assessment – Purpose and Goals
 - Summative assessment is important because we need to know if students have met the expectations of our standards each year. This way, we don't let them fall too far behind before intervening. Parents need to know this information.
 - Summative assessment is useful when we do actually USE the information. Summative assessments should be used for school improvement planning, to inform instruction and support students.

- Summative assessment helps to ensure equitable access to state standards. The disaggregated information helps us to know if we are serving all student subgroups well and maintaining universally high expectations.
- Summative assessment should not be used as the primary source of information for student progress monitoring or daily instructional planning – formative assessment is the correct tool for those purposes.
- Formative Assessment – Purpose and Goals
 - Formative assessment provides diagnostic information about student skill gaps. This is the primary source of data for student progress monitoring.
 - Formative assessment helps teachers and students know where they are in terms of standards mastery. It helps build confidence for the summative assessment.
 - Formative assessment must align with the expectations of the standards and the rigor of the summative assessment.
 - We should balance the emphasis on formative and summative assessment. They are equally important.
 - Formative assessments drive teacher collaboration when they are created by teachers. These common assessments provide actionable information about the needs of students and planning instruction.
 - We should use formative assessment to help support the highest-achieving students, such that we continue to challenge them.
 - Formative assessments must include questions that are valid and reliable in terms of alignment to the standards and reflecting the rigor of the state assessment. Districts need support with item banks to create valid and reliable formative assessments.
- Test Preparation – Culture and Goals
 - The best test preparation is having teachers with an absolute and complete understanding of the standards and how they will be assessed.
 - Transparency in the test design – including blue prints and question format – helps teachers and students know what is expected of them to demonstrate mastery of the standards.
 - Clear alignment between the standards and the test will ensure that we test what is taught.
 - Test preparation is every day; it is good instruction.
 - Test preparation is not “drill and kill.” Students retain information through spiraling skills all year long and teaching with as much intensity on the last day as the first day.
 - Classroom assessments should be cumulative in nature, including not only questions from new units, but also questions from prior learning.

Assessment Task Force Meeting #3

June 16, 2015

- Commissioner McQueen welcomes, group introductions
- Commissioner McQueen overviews the state reports
 - The audience of the report is general in nature: legislature, district and state leadership, principals
 - Primary areas of focus for other state's recommendations?
 - Colorado eliminated a lot of tests, accountability was flexible (Took away responsible for ELL students; gave more flexibility on subgroups)
 - If we reduce accountability, when will it slip?
 - Technology readiness was a heavy theme among all the reports
 - Professional development report
 - Time demands of assessment
 - There was a problem stated, but the solution was not clear
 - Reading readiness – skill specific, fully aligned K-2 helps you be ready for third grade assessment
 - What did you like? Dislike?
 - Liked specific recommendations in the Colorado report
 - Liked Colorado specifically addressed ELL students
 - Liked tensions mentioned in the Colorado report
 - What should our report include?
 - Reflect concerns with specificity
 - Address accountability model – Will address during July meeting
 - Is our TVAAS restrictive?
 - Tennessee is very unique in our value-added model
 - Including next steps
 - We should address special education students
 - If we explore opt-out, we need to address the current policy, the state law on the effect on grades, etc.
 - This may have been a bigger problem because of Common Core perceptions; some of the opt-out conversation may settle now that we are having the standards review
 - Statement about culture of testing
 - How are assessments good for students?
 - How do we communicate this report when it gets published?
 - How and what way would this group make recommendations on RTI?
 - Recommendations must align with principals

- Quick Scores, Nakia Towns –
 - See handouts provided for detailed information
 - 2011 – Legislation passed to include TCAP scores in final grades
 - People came to think that proficiency was tied to quick scores; they are not – quick scores are taken from raw scores
 - Scale scores are where the cut points lie
 - All accountability scores come from the scale scores
 - Why did we choose to change the methodology on the last year of this test?
 - We were going to end TCAP and move to PARCC in 2013-2014, so we could not use the interval methodology. (You have to have initial data to use interval methodology.) Therefore, the department decided to use the method that is used in high school for the new PARCC test which doesn't need initial data. Legislation then moved us back to TCAP, and the decision to change the quick score methodology for 3-8 was not revisited.
 - What do we tell parents about the quick score? How do we change the perception of quick scores? How do we help teachers, parents, and students understand?
 - Takeaway: this is a continuing conversation; the department needs feedback from stakeholders about quick scores
- Test Development, Emily Freitag
 - Test design terms can be found in your PowerPoint
 - All starts with the standards
 - Design a test on a macro level- blueprints
 - Test items get written by vendor
 - Then the test questions get reviewed by multiple teams of teachers looking for sensitivity, accessibility, etc.
 - Items must be field tested (this is a “trial run”)
 - Choose best items and create a form (an operational test)
 - Test Components – see PowerPoint for full description
 - About 70% of the test every year
 - Unique operational items – reflected in students' scores – these can be released after year one of TNReady
 - However, if we do this, we will constantly have to create new questions, and there is a budget implication for this
 - Paying teachers to review, vendors to write, etc.
 - Linked items – reflected in students' scores (we cannot release these if we are going to use them again)
 - 20-40% of the test every year
 - These help us compare student performance from year to year

- Context from experience that might infirm the state decision around TNReady, Aspire, EXPLORE, and PLAN?
 - Best report that parents receive (EXPLORE test)
 - If this can be replicated in TNReady, I don't know if I need it
 - I am losing the quick data that we use for high school placement
 - Use it a lot for eighth counseling/ninth grade placement
 - We give students individual ACT goals derived from this test
 - Could use the SASS predictions; it's just easier to communicate using the EXPLORE test
 - Identifies students that may need some more help in eighth grade before the transition to high school
 - TNReady and EXPLORE need to be comparable
 - TNReady must be aligned; it is crucial
 - Could we provide ACT predications on third graders?
 - Would be relatively easy to do; states have done this in other ways
 - Anxiety for parents
 - You don't want the di cast – “It's all about goal setting.”
 - Does ACT prep help?
 - The state is gathering this data
 - Data in Knox County shows that tutoring outside of class has a significant improvement in their score
 - Quality tutoring is key
- Perceptions about PLAN and EXPLORE?
 - Easily understood by students, parents (tells story better than state tests)
- K-2, Emily Freitag
 - No state mandated test can be conducted earlier than grade three – Legislation
 - Offered SAT 10 for the last few years
 - Pearson product that is being phased out
 - Measures student mastery in early grades
 - Norm referenced test (compares performance to other students, not students' knowledge about the standards)
 - RTI – interventional model (for details, look at PowerPoint slides)
 - Old model:
 - When students were struggling, they got referred to special education
 - “Wait to fail model”
 - New model:
 - Screen students in reading and math in skills (not standards)
 - Helps capture struggling students earlier
 - Help close skill gaps

- Universal screening has become a part of the K-2 experience, so what assessments can be used in those grade levels?
 - Pretest, posttest model?
 - Criterion referenced test?
 - Would continue to be optional
 - Questions?
 - Would this be online or paper/pencil?
 - We are thinking about procuring both, and you would be able to choose the direction you want to go
 - How did we land on the pretest/posttest design?
 - Would allow for a value-add score
 - Becomes very actionable
 - Additional context would the group like to share about your perspective on K-2 assessments?
 - A lot of communities don't like SAT 10, so you would have to be sure you communicated the differences
 - Seems important to the state
 - Universal screener data does not make sense to make a value-added score; they are designed so that you cross a threshold
 - Districts choose the universal screener that they want
- Report Out:
 - Clarity – Must get reports correct
 - Before spending on retests, make sure reports are right
 - Perhaps a Parent Advisory Group to make sure reports are understandable
 - Find another way to get exploratory career information
 - Assessment logistics, technology integration
 - Align state funding with technology

Emily Freitag's Group

High level:

- Summative assessments are too late for intervention – should not use that word
- Reporting – extremely important, how can we build that out?
- Does all students tested include K-2?
- I want to pull out the logistics piece – in our focus groups scheduling and technology are the biggest things coming out as areas in need of support. Testing logistics – it probably needs to be expanded. Testing delivery decisions, while maintaining the security of the assessment. Separate bullet – testing security?

- General assessment literacy – helping everyone understand the differentiated purposes. What TNReady can and cannot do – be clear about the purpose. Without overselling the use.
- Formative recommendations – feel like there is a lot here – district flexibility is important but should we build out more on district support? Support at the district and school level. There is a lot of clarity needed on the formative assessments with purposes – all mixed up. Multiple instructional practices.
- High expectations for students about use of technology and importance of going in that direction?
- Transparency around all of the pieces of assessment and all the elements. There may be tensions between transparency and security but we can be much more transparent.
- Test prep side – resetting the conversation but making sure prep for technology does not replace prep for bubble questions. How to refocus on teaching the skills.
- Create a statement at the header that is about this is why we do testing – preamble. Purposes of testing.
- Need to add something on the opt-out piece.
- Focus on growth.

Recommendations:

- Test quality – we must invest in ensuring strong quality in the design to be sure that the test actually measures the skills that students need and regularly review the quality of our assessments (perhaps alongside standards reviews?)
 - Perhaps group the recommendations structurally – state, district, legislature:
 - Legislature: Annually release the unique operational items and afford the budget to do so and state results on each item.
 - State: Annually release blueprints and test specifications and how reports are calculated.
-
- ACT: Attractive to streamline the tests however what is it going to cost to get TNReady to give you those kind of reports – more, the same or less than ASPIRE?
 - Only worth dropping if we can get that information in other ways
 - Retest? Focus first on using money to provide better reports. Will you detract from initial performance or will it just be short term gain?
 - Is the retest for everyone?
 - Are you negating the benefit of second tests by taking away the first?
 - Do think the taking a second time is important, especially for low ses students
 - Is there that much value to those kids?
 - Has to be a plan in place to improve the score – not just retake
 - I think it should be for everyone – all can improve
 - People use explore for career exploration. – that is valuable. – we would have to have an alternate approach for this. (Will aspire even continue to offer this?)
 - If you can communicate clearly and provide good reports that break it down to parents simply, then there is value in streamlining.

K-2:

- There needs to be a lot more definition about purpose of assessments and the difference between assessment and screening
- Clarity around purpose.
- What is the purpose of screening versus assessments
- Been an outcry for K-2 assessments – is it value added or students?
- Not tied to district accountability
- I am good with K-2
- Pretest/posttest teachers like that
- Challenge to pretest posttest – not on standards that have been taught
- Reports pretest become quite important and must be thought through carefully ahead of the RFP
- Norming assessments K-2 gives them information about norming nationally
- Emphasize parent communications of screeners performance data
- If tests are given, the results must be shared with parents
- Form a parent advisory group to look at the reports across all assessment
- How screeners and skills align to standards

Nakia Towns' Group:

- **Annual/Summative/Standardized Assessment**
 - State testing should be properly aligned to state standards.
 - This is KEY!!! If this aligned, then local and classroom will follow. Not as much benchmarking required. Would trust common formative assessment.
 - Summative assessment should show *all* students' culmination of understanding, strengths, and areas for improvement to stakeholders.
 - Summative assessments help us intervene in a timelier manner. It also helps ensure equitable access to state standards. The disaggregated information helps us know if we are serving all student subgroups well and maintaining universally high expectations.
 - Grade-level annual assessments show progress toward postsecondary goals. They also prepare students for postsecondary entry.
 - Clearly communicate track progress at each grade level towards and benchmark at terminal point e.g. ACT. Clarify ACT and postsecondary readiness target and backwards mapping.
 - State assessments should help bridge expectation and communication gaps among teachers, schools, districts, and the state.

- Identifying gaps that are created for students by standards transition? Skill mapping to clarify skills.
 - Reports should be clear, readily understood by the district user, and provide timely feedback.
 - What is timely?
 - Annual assessments point to level of rigor needed in the process of teaching and learning, can define adult responsibility, and can help us benchmark our success against other states and countries.
 - The community needs to know as much about the test as possible. The state should release some percentage of questions every year to create clarity, transparency, and trust.
 - It is required by federal and state law.
 - Need to connect this to messaging for OPT out parents.
 - Assessment is part of the instructional process and the cycle of learning. It is not a punishment. To foster this attitude, we must remain focused on the bigger goal of preparing students lives, not the test as the outcome.
 - Summative assessments should be used for school improvement planning, to inform instruction and support students, not as the primary source of information for student progress monitoring or daily instructional planning.
 - Clarifying PD connection for teachers.
 - We have to invest in making sure this is a priority.
- **Interim/Benchmark/Formative/District Assessment**
 - Formative assessment provides diagnostic information about student skill gaps. This is the primary source of data for student progress monitoring.
 - Formative assessments that are teacher-led best inform instructional changes and these are preferred to general, large-scale possibly unaligned off-the-shelf assessments.
 - Formative assessments should have clarity and be used for specific decision-making and next steps so students can effectively build future skills.
 - KEY POINT. Can we expand uses of required tests?
 - Students deserve to receive timely feedback about their progress.
 - Each formative assessment should have a clearly communicated purpose that all stakeholders understand.
 - Clarify each stakeholder group.
 - KEY POINT. For all assessments. Eliminate redundant assessments, i.e. those for the same purpose.
 - Formative assessments should be used to measure mastery on standards that have been taught and must align with the expectations of the standards and the rigor of the summative assessment. They should build confidence for the summative assessment.
 - Formative assessments should be designed in a variety of forms, not just paper-and-pencil (look at academic and soft skills). Intentional cross-curricular and multi-purpose formative assessments are encouraged.
 - This is critical to TNReady, informational text, and connections between content areas.

- Districts should have flexibility to design their own programs but should have support when making tests. This could be accomplished by having access to item banks that could help create valid and reliable formative assessments and offering professional development about how to write a great test question.
 - A review of grading practices should be conducted.
 - How do we manage through a transition for student grading, given new proficiency. How do you translate standards based grading to the 9-12 A-F grading scale? How do we communicate to parents? What are the implications for HOPE? College admissions, etc.
 - We should balance the emphasis on formative and summative assessment. They are equally important.
 - Formative assessments drive teacher collaboration when they are created by teachers. These common assessments provide actionable information about the needs of students and planning instruction.
 - We should use formative assessment to help support the highest-achieving students, such that we continue to challenge them.
 - Formative assessment, though a best practice, are entirely a district option.
- **Test Preparation**
 - Teaching the standards should be the best preparation – the best test preparation is an absolute understanding of the standards and how they will be assessed.
 - Test preparation is every day; it is good instruction.
 - Change this order to emphasize instruction not test prep.
 - Testing strategies should be part of the regular teaching and learning cycle. We should have unrelenting high quality instruction.
 - Classroom assessments should be cumulative in nature, including not only questions from new units, but also questions from prior learning.
 - Teachers should take care not to pass on stress.
 - Maintaining professionalism in terms of communicating with students regarding importance of test. Careful language in communicating.
 - Technology must be integrated in the teaching and learning cycle on a regular basis. This is preparing students for the digital age.
 - KEY POINT. Technology process.
 - Students should have the opportunity to develop familiarity with the test platform in a natural way and in instruction.
 - Transparency in the test design – including blue prints and question format – helps teachers and students know what is expected of them to demonstrate mastery of the standards.
 - The state, districts and schools must work toward daily technology usage for every student.
 - Clear alignment between the standards and the test will ensure that we test what is taught.
 - **Test Logistics**
 - The test should provide maximum flexibility in all testing decisions.

- The state should provide more test support with opportunities for input from those on receiving end of tests.
 - Computer lab? Working through instructional lab. Using technology daily. Integrating technology on an everyday. High school? Type mathematically?
 - Testing decisions should find proper balance between testing time and instruction.
 - Tests should be short enough to fit in a normal class schedule but students should not be rushed.
- Writing/typing skills with 90 minute administration (processing information)
 - Practice test around technology. MICA integration in classrooms.

K-2 Implications

- Considerations for individual accountability (growth portfolio versus value-added)
- Online assessment versus paper?
- What is best for Kindergarten? RTI??? Diebels? Pathdriver?

ACT Implications

- EXPLORE seems redundant. We have enough information? Getting information for transition. Not utilized by teachers. Under old standards/cut scores it was useful. Not as much now.
- ASPIRE seems to not be as valid without adopting full series.
- National performance information – standards setting with external benchmarks and other states using item banks.
- AP rigor versus regular high school courses.

Commissioner McQueen's Group:

Recommendations

- At the state or district level, if we administer X assessment, then we will know why for the purpose of Z. Parents need to know “why” and information should be posted on the state and/or district websites for clarification.
- We need to explicitly educate future and new teachers about assessment for learning. Teacher preparation should include a specific curriculum or module on assessment with specifics to Tennessee. All new teachers need specific training as well with embedded PD around this area of the TEAM rubric.
- Use principle around % of test questions.
- Aligned assessment, clear reports – assessing for learning (strong alignment to postsecondary).
- Transition away from Aspire (Explore and Plan) and begin to define better reports and timeliness of reporting.
- Create 2nd grade test as alternative to SAT10. Create K-2 RFP for pre/post test.
- Work keys?

Assessment Task Force Meeting #4

- Commissioner McQueen welcomes and goes over task force goals (in PowerPoint)
- Commissioner McQueen, 2015 TCAP Statewide Results:
 - 3-8 students gained in math and science, but there was a decrease in RLA
 - At every single grade level there an increase in math scores
 - Student growth in math has eclipsed reading in the last five years
 - 3-8 RLA distribution is flat (Aren't moving students from basic to proficient, etc.)
 - High school students improved significantly in ELA; largest jump in English III
 - Black, Hispanic, Native American, economically disadvantaged all made gains
 - This signifies that the achievement gap is closing
 - English Language Learners and students with disabilities did not make gains
 - Students with disabilities moved off the modified TCAP test and took the TCAP this year (+16,000 students)
 - How do you account for the gap between high school and 3-8?
 - Different test, different standards
 - We have to analyze what we are missing in 3-8
 - There were actually slight gains in 7th and 8th grade RLA that were not seen in 3-6, so what is happening there?
 - The way teachers teach has drastically changed in last three years, but the test has remained the same. That could mean growth with TNReady.
 - You can exempt 1% of students from the test, and there is a portfolio option
 - As a state, Tennessee is slightly over the 1%, and the redistribution has not occurred yet. (It occurs, not at the child level, but the district level.)
 - The amount of students that districts can exempt varies as long as the average at the state level is 1% (i.e. some districts may exempt more than others)
- Alyssa Van Camp, SCORE study
 - Completed the focus groups, analyzed the data
 - There is a survey recap in your PowerPoint
 - Focus group was small, so take results seriously but look at it with a critical eye
 - Top four codes by sub-theme: (All are in PowerPoint)
 - Challenges with assessments – by far the largest:
 - Negative experiences, comments, and struggles with assessment
 - A lot of time being spent on assessment
 - Limitations of current assessments
 - Special populations
 - Student anxiety

- There is so much variability from district to district in the way that assessments are being administered and used
 - While intentions are all the same, people are all over the place on what the best practice is
- There is more teacher buy-in when teachers lead changes. Teachers need to know where students are, and formative assessments are important. Isn't it better for teachers to create them because they know what they have thought?
- Teacher Educator Survey: Preliminary Results, Nate Schwartz
 - Survey was e-mailed to all teachers and administrators
 - School and district results will be released in August
 - About 62% of teachers believe they spend too much time helping students prepare for statewide exams; Almost 70% of teachers believe students spend too much time taking statewide standardized exams.
 - This is just a "how you feel" question
 - All across K-12 this is a concern
 - Every little test in high school disrupts the whole school because there are often tests that span across grade level. (This disrupts multiple classes.) Also, teachers cannot administer tests in their subject area in some districts; teachers can never administer a test to their class in high school.
 - We don't want the perception that we shut down schools to administer assessments
 - Don't forget that the survey was given during the time that many assessments were being given!
 - Almost 75% of teachers who have tested subjects believe that they spend too much instructional time preparing students for statewide exams and that students spend too much time testing
 - Maybe students are not taking too many tests; maybe students are out of the classroom too much because of testing: is testing the problem or is the loss of instructional the bigger problem for teachers? Scheduling issue?
 - More teachers understand how to use data from assessments (69 % → 76.6%)
 - More teachers are being data-driven
- TNReady Communication, Ashley Ball
 - Created a parent guide, roughly 20 pages
 - FAQs, infographics, parent checklist to prepare for TNReady
 - Side by side comparison of TCAP/TNReady questions
 - Testing times are comparable, and the guide shows that (Because TNReady has a Part 1/Part 2)

- Accountability Overview, Nakia Towns
 - Scheduling and Logistics Task Force began this morning
 - Why do we have accountability systems? What is the history?
 - Political framework
 - Return on investment (Education one of TN's biggest investments)
 - Economic development
 - Equity (All groups should reap individual rewards from education)
 - Theoretical framework
 - Principal-Agent Relationship
 - Agency relationship: you depend on others to reach goals
 - Must ensure alignment
 - Ex. Focus on STEM from state legislature (principal), but the teachers (agents) want to be sure the whole child is developing – How does everyone meet their goals?
 - Monitoring (Ex. Performance management)
 - Incentives (Ex. Pay/promotion based on performance)
 - Sanctions (Ex. Demotion/termination)
 - Accountability model based on robust, intentional way to foster systematic change
 - Annual assessment began in Tennessee in the 1992 (TVAAS) Not in 2001 with No Child Left Behind
 - 2005-2006: EOCs began
 - 2011: Introduced student grades that incorporated TCAP into grades 3-8 (tried to bring students into the accountability model)
- What are the ESEA waiver changes that are coming?
 - 2012: Earned approval for its first ESEA flexibility waiver from the federal government DOE and developed own state accountability system
 - Waiver expires at the end of 2015, and we resubmitted our waiver application (waiting to hear)
 - System proposed in the waiver application
 - Recognizes incremental gains by rewarding partial credit when districts don't meet targets
 - Will work through the TNReady transition
 - Proposed accountability system overview
 - Step 1: Minimum performance gate (In need of improvement)
 - Then every district gets recognized for progress they are making
 - All key changes in proposed system are located in PowerPoint (slide 13)
- Three types of schools:
 - Reward schools

- Performance: top 5% based on achievement
 - Progress: Top 5% of school in TN based on TVAAS growth
- Focus schools
 - Largest achievement gaps between groups in the state OR low performance for subgroup based on achievement or graduation rate
- Priority Schools
 - Lowest performing 5% of schools based on achievement
 - Schools can be exempt if meeting annual goals
- Achievement School District
 - Focused on moving out students in schools that are in the bottom 5% of schools to a state run district (a turnaround district)
 - Authorizes charters
 - Direct run schools (run by the state)
 - What happens if schools get out of the bottom five percent?
 - Schools will be returned to the district once a plan of continuous improvement and leadership are in place (ensures that improvement continues)
- Teacher Evaluation
 - Five levels that range from below expectations to about expectations
 - What makes up the score?
 - Classroom observations
 - School or district-wide and/or individual growth data (TVAAS or portfolio)
 - Other student achievement measures
 - Student survey data (optional – determined by districts)
- Tennessee Teaching Evaluation Enhancement Act
 - Adjusts the current weighing of student growth data in a teacher’s evaluation during the transition to TNReady
 - More detailed information can be access in PowerPoint (slide 21)
 - Next year teachers will be able to choose their one year composite (using only TNReady scores) or their three year composite (using both TCAP and TNReady scores), whichever is higher, for their teacher evaluation
- Public reporting
 - School, district, and state level reporting is online
 - Individual teacher evaluation data (including TVAAS) is protected from public release
- Overall...
 - Accountability model is about progress
 - “What gets measured, gets taught” – we need to help teachers focus
 - Gap in district level accountability; state helping support this closure

Feedback on the Recommendations from Nakia's Group

- Timelines for release blueprints & test items - No later than mid-July. Item analysis - by the June 15 & June 30 with TVAAS. Getting that feedback is important
- BEP Review Committee to potentially address funding for item release and resulting development?
- Communication on testing windows due for posting to district websites by July 31. How do you communicate about the individual timing, as the testing windows are longer but not all students are involved? Student level examples...
- District TNReady representative. Who is the expert in the district leading parent information sessions? Communications specific to the expectations of schools and districts.
- Purpose of accountability? Use the information that we provided. Create an FAQ. Arm Schoolboard members with their own "TNReady Guidebook." Provide examples of what other school boards have done to educate community about assessment. Provide private-company CEOs with continuing education about education reform. Use TSBA connections. Keeping school boards and business leaders up to date. What is happening and WHY it is happening? FAQs.
- Pre-K/K/1 grade Portfolio. Important to have individual accountability measures based on 2nd grade assessment. Success in the third grade is too important to not have a measure of who is on track at the end of second grade. Can we make district participation in the 2nd grade assessment mandatory, but make generating TVAAS for 2nd grade teachers optional? What would be the 2nd grade growth measure if districts did not want to do 2nd grade TVAAS?
- Should we assess social studies and science in 3rd grade? Some of the social studies standards are "one and done" in a particular grade level. Will people teach the standards without the assessment?
- ACT support. If we do retake, then we can do more remediation. What is the purpose of the retake? Further analysis needed on why students are not taking the ACT a second time in some schools and districts? We need to explicitly link this to repurposing of funds gained from eliminating EXPLORE and PLAN. What about family responsibility for paying for second test? Financial Need? Perhaps we need to require that students commit to additional coursework to support the state investment in the student taking a

second time. Will students take the first test seriously if they know a second test is free?

- Need more communication with teacher prep programs about assessment and TVAAS. Need teachers who are better prepared and have a better of understanding of professional expectations.
- What's missing?
 - Technology SKILL not just technology readiness. Typing/keyboarding. Starting early in 1st/2nd grade with intentionality.
 - Technology investment. Where are the funds coming from?
 - What about the broadband coverage? Some rural areas don't have infrastructure.
 - Special populations students. Recommendation for accommodations and modifications based on IEP.
 - Gifted children. A specific statement?
 - College and career standard defined at 11th grade and backwards mapping to the lower grades. Alignment throughout! Rigor of expectations cleanly aligned.
 - RTI recommendation/universal screener. Diverse learners - personalization.
 - Career and Technical education. HS Redesign. WORKFORCE Readiness.
- We still haven't reached the national average for NAEP. As well as we are doing, we aren't there yet. Keep pushing!
 - Alignment and giving a clear map for success. Roadmap for success in 2009 from SCORE. Mirror that was shown about the performance for our kids.
- What is the BIG PICTURE??? Assessment is not a punishment or gotcha! A bomb that is being used against them. The final report from the taskforce needs to start with the big picture.
- Leverage advocacy through Teacher Advisory Council (Teachers Cabinet) and SCORE Teacher fellowship.

Feedback on Recommendations from Nicole and Vicki's Group

- Overall
 - Add overarching messages about the “why” of different assessments
 - Include the best preparation is quality instruction every day
 - Seems to be missing mention to address the amount of testing

- Summative Standardized Assessments
 - Should #8: be clearer about what is meant by transparent: maybe reword as “Provide great transparency to question types, subjects, and reporting to standard/cluster level

- Formative Assessments
 - Should #4: Group was very positive about this principle but need technical assistance on effective assessment design
 - Universal screeners/progress monitoring – not sure that this should be included here but seems to be missing generally

- Test Prep/Logistics
 - Districts should #4: recommend how and support teachers and leaders; support for students; should we consider the role of counselors during testing time (e.g. as testing coordinators they are not available to support students)
 - Add a principle “Thoughtfully schedule assessments to prioritize normal instruction and minimize disruption.”
 - Some conversation about whether there should be a bucket for what schools should do or if these are included in the district bucket

- Recommendations
 - Seem to be missing recommendations to reduce testing (time and number) – at least not very clear (#8 does reduce number)- and a gap for grade 1 teachers
 - Chairman Brooks thinks that items 1, 5, 8, and 9 will have support from legislators, thinks legislature would be willing to allocate some funds to supporting #1
 - #3, in addition to guidance and support (e.g. sample schedules), support should include process guide to help other districts replicate the work
 - #4: clarify which reports – recommend including parent advisory group on school, district and student reports
 - #5: split formative and summative; include why for each assessment type – summative, formative, universal screener, progress monitoring
 - #7: 2nd grade assessment should remain a district option – think this is the intent but needs to be clearer
 - Add recommendation
 - to provide item bank for district assessment use
 - Develop assessment literacy for current teachers, in addition to pre-service teacher, to include test anxiety, criterion vs. norm-referenced assessment, support students with test anxiety

- Chairman Brooks was positive about the principles and recommendations. His comments were that some of these issues as addressed will be received very positively by the legislature. Specifically, he mentioned #1, #5 and #8. He feels positive about funding

necessary pieces to implement. He also stated that the public will like #7. Public needs to understand clearly how we are reducing testing time. He also wants everything possible to be released, within reason for test security, linkage and cost. He understands there may be a challenge around communicating why all test items cannot be released. This should be a point for education – for the public and for the legislature.

- **Larger issue** – perhaps an introductory piece that outlines the overarching purpose of assessment and defines and delineates among summative, formative, universal screener and progress monitors.
- There was a question about how ACT would be counted for accountability.
- **General comments:**
 - There is no recommendation around test prep. It is in the principles. Some mention of solid daily instruction being the best test prep could go into an introductory piece (as mentioned above).
 - There were concerns around test anxiety. Some thoughts: testing coordination should be managed by someone other than the school counselor (which frees that person up to deal with anxious students – and teachers). There was a feeling that there should be some recommendation around testing culture – about minimizing test anxiety. (This is in the principles but not the recommendations.) Some comments: leadership matters with this issue; how can we support teachers; develop an understanding of what test anxiety is – what causes it and how to allay it.
 - With regard to reports, there was a feeling that reporting should be called out by recipient group – so transparency and reporting is important for teachers, but parents also need good information, and we need to provide for them the most detailed look at how their student did as possible.
 - Banding subgroups together in reporting needs to be carefully communicated so it is not perceived that we are masking results.
 - Include something about the state providing technical support and/or provision of items for formative assessments so districts can develop their own formative benchmarks.
- **Specific comments with regard to recommendations:**
 - #3 Some statement about thoughtful scheduling that minimizes disruption is needed; some discussion about specifics (I don't think these are appropriate for

recommendations) but questions about supports to districts (sample schedules, process guides)

- #4 The group thought calling out recipient groups for reports would be helpful: “provide clear reports, state, district, school, teacher, parent”

- #5 Should this recommendation be one point or two? One for summative and one for formative? Also expressed interest in including information regarding addition of other assessments besides 2nd grade: portfolios for preK, K and 1? When communicating the “why” teachers should be added to parents and students.

Feedback on Recommendations from Commissioner McQueen’s Group

- Report:
 - Add a definition section to the report to define terms
 - Note district variability in how assessment is being handled and seek to support district leaders
 - Specify theory of action for assessment – why assess?
 - Connect analysis section to what should be done next year – continued process of improvement
 - Acknowledge non-tested grades and subjects

- Recommendations:
 - Explain why all test items can’t be released
 - Note need for more transparency
 - State should create document with guidance for leadership around patterns of behavior that would relate to nonrenewal of teachers
 - Items banks are needed

Assessment Task Force Meeting #5

August 24, 2015

- Stephen Smith welcomes and goes over task force goals (in PowerPoint) and agenda
- Update on post-secondary group (UT System and TBR)
 - Explores TNReady and post-secondary alignment
 - Talking about cut score institutions will be using to determine readiness
 - Both systems are in the process of choosing faculty to serve
- Stephen reviews progress to date

- **Nakia Towns, TNReady Score Reports**
 - One of the major drawbacks we were facing with TCAP: Was the information actionable and accessible? Did students and parents know what this means and what to do with it? We want TNReady reports to answer these questions:
 - How is my student performing according to standards?
 - Areas of strength? Improvement?
 - How is my student performing relative to the school? District? State?
 - What can parents, students, and educators do now?
 - Many people found the ACT EXPLORE report very useful, and we have started mocking up the new report design to look more like this report
 - What are we considering?
 - Single subject vs. portfolio
 - Performance level names/descriptors (What do these actually mean?)
 - Score scale ranges
 - Color scheme, orientation

- **Whole group discussion on TNReady reports:**
 - I like having the scale scores and that they continue overtime (Comparative information)
 - More immediate visuals – colorful
 - Landscape, large font, and simple layout were appreciated
 - Have the number of performance level and the name of that level together on a chart together; do not separate them; show where student falls on scale
 - We liked “On grade level, above grade level, etc.”
 - Terminology of “beyond grade level” does not feel right because that is not what we are testing them on
 - Separate parent and teacher reports should be considered
 - Less suggestions may be more helpful to parents (more next steps for teachers)

- The content strand (number of questions, how many student got correct); this may be more teacher friendly, but I really like that
 - Adding the number of omitted (skipped or not finished)
- Scale of the report:
 - TNReady will have a 200 scale score
 - Should start at 0 because it seems like a basic start line
 - Who prints it? If it is at the district, there is no use doing color
 - Using the same color scheme that TVAAS using
 - If grayscale is going to be used, be sure that it still makes sense
- **Nakia's group discussion on TNReady reports:**
 - Multiple Subject - Student represented with a different symbols
 - Color Scheme - red, yellow, green, blue
 - Label - Ready for acceleration. (vs. Advanced). No. Accelerated.
 - Below-grade level or Above-grade level
 - Distinguished
 - Too busy on portfolio
 - Portfolio on individual students for Teachers and single subjects for parents?
 - Average Scale Score seems to provide a better view of the comparison of state/district/school/student
 - Checklist of strengths and weaknesses is good for parents
 - Next steps - 2-4 for parents (maybe more for educators)
 - Prefer "grade level" language for performance levels
 - Next steps - Advanced (consider subject)
 - Use of color is good
- **Scheduling and Logistics Advisory Council**
 - TNReady has more flexible administration (longer testing window)
 - They created models that allowed instruction to continue while tests are being administrated, and we have provided these to testing coordinators (The council created what they believed would be the best for their school and the exemplar schedules were pulled and distributed.)
 - Nakia went through the examples, and they can be found in the PowerPoint
 - Even if the testing window is 10-12 school days, no one student or teacher will be tied up for the entire time
 - These schedules are built to continue offering instruction; there is no "school shut down" (This can now be done because schools are not forced to do it in a certain five days.)
 - Tablets are acceptable to take TNReady (The technology does not require a keyboard, but the state has said that is best practice.)

- Wireless connection is required; the test writes to a server every 90 seconds, so if something catastrophic happened with the technology, the most a student would lose is 90 seconds.
 - How long are EOCs?
 - ELA Part 1 – 90 minutes
 - Chemistry/biology – 75 minutes
 - Students finish ahead of time, what happens? LEA decision
 - Tools to help administration with testing: decision tree, calendar template, etc.
 - Group will continue work for rest of the school year
 - “Break MIST Day” – we want to break it when it does not matter (fire drill)
 - Test server capacity and locally they can check bandwidth/capacity
 - There is no prohibition against students bringing their own technology, but it requires adding a program (it is not web-based)
- **Whole group feedback on the ATF report, Stephen Smith**
 - Recommendation 1 Feedback
 - What are other means to communicate to parents other than the website? (Some parents do not have access to the internet.)
 - A suggestion made by a principal is that the state sends a parent to letters
 - Letter in the school toolkit that could districts could adapt?
 - Recommendation 2 Feedback – There is a price tag
 - Recommendation 3 Feedback - Don’t “calculate” reports; work on the wording
 - Recommendation 4 Feedback – None
 - Recommendation 5 Feedback –
 - Should include a recommended practice of sharing with parents/students
 - Recommendation 6 Feedback – None
 - Recommendation 7 Feedback – None
 - Recommendation 8 Feedback –
 - Just first grade? No, but specifically first grade for this recommendation
 - Portfolios are not to assess children; we need to be clear on the purpose
 - Language should be changed to reflect that these are district/locally driven
 - What about the current portfolio options?
 - Recommendation 9 Feedback –
 - Is this going to be required? This feels like it is against the task force goals
 - The test will not “improve literacy,” so we need to change the wording
 - We want a test that is aligned; we never discussed if it would be required
 - Helps address the “imbalance” at the early grades; where are you going to put your best teachers with our current model?
 - “The word “require” will cause all sorts of legislative issues.”
 - Does this actually help? If so, what data do we have showing that it helps?

- Portfolio would not continue into second grade
 - Pre-K, K, and first grade have portfolios to measure growth
 - Second and third grade teachers growth measures are measured on school-wide growth
 - SAT 10 for K and 1 is done in spring 2016 (assessments being removed)
 - We need to try get superintendents on board to avoid legislative issues
 - Recommendation 10 Feedback –
 - Include validation of TNReady by higher-education community (it is progress monitoring)
 - What does “other methods” mean?
 - Be explicit in saying what other methods is (career inventories)
 - This is a “rambling” recommendation; it is too early
 - Eliminating state-mandated testing
 - Call out that we are not replacing SAT 10, EXPLORE, or PLAN
 - We need to call out that there is an assessment shift this year, so we need to re-examine this again next year to see if things shake out the way they are supposed
 - Currently with the charts online, it looks like there is testing every day. We need to communicate that even though the testing windows are longer, there is actually less testing
 - There is a lot of assessment on the teacher side; things have been added (screeners), and we need to sensitive to that
 - Recommendation 11 Feedback
 - Does this legislatively have to be acted on? Yes, it must be reflected in the budget
 - You should not limit it to a low score
 - Require action on the part of the student to qualify for the retake
 - Recommendation 12 –
 - While students are in college, they get a class on assessment; would this be better as a workshop for first-year teachers?
 - Would a class be better then have a follow-up workshop?
 - I think having it during a prep program is important so you can apply it in your practicum, which is important.
 - What about training for our current teachers?
- We need to address the greater question of how much testing is too much testing; this is a further analysis question; we need to put it on the table so people do not think it has been overlooked (It is not an easy solution to say “cut this test”)
 - Recommendation 13 – None
 - Recommendation 14 –
 - Has there been a proposal to remedy this situation?

- This is a “thorny conundrum”
 - Is this worthy as a standalone recommendation? Should it be combined?
 - Should not say that counselors cannot help with assessment, but that they should be available to children during assessment
 - Recommendation 15 –
 - Should 14 and 15 be combined?
 - Is 14 a next step for recommendation 15?
- What are the next steps for these recommendations?
 - We got to here from previous meetings; made several revisions and going to take the feedback from today
 - We will need collaboration with other entities to get these implemented
 - Call on other entities in the report that are required to implement things so it is clear who is needed to help
- **Nakia’s group feedback on the ATF report:**
 - **Definitions:** Specify that they are referencing this report for summative assessments
 - **Summative:** Add a synopsis of the work that we actually did
 - **Formative**
 - Special education, if students are IEP what that the portfolio. Teacher led - very individualized for students. Scoring the students. Behavioral goals. The IEP guides assessment practice in referencing the state. Accommodations will be made.
 - **Testing & Logistics**
 - Reliance on labs versus access to individual computer. Technology themes?
 - Technology access?
 - **Recommendations**
 - Opt- out? Not specifically mentioned or addressed.
 - What about EXPLORE/PLAN elimination and parents who want that tracking toward ACT? There may be resistance to that? What would let go of?
 - #1 and #6 should specifically reference parents
 - Add to #5 - Communicate about TNReady with Higher Education vetting TNReady and validating it.
 - #7 should be earlier in the earlier in the list as it defines the purpose of formative assessment
 - Communicating about test complexity? How does information get out beyond the internet? A letter to parents from the state? About what?

- What about writing? Writing on demand? Responding to text?
- What do teachers do with the results of TNReady? Actual conferences around TCAP results. Consistent practice around conferences with teachers and students?
- Career - Work Keys Under CTE??? Capturing in the recommendations.
- Recommendation Portfolio - reference teacher evaluation and student growth measures. Student growth measure for teacher evaluation. These also have districts and schools
- 2nd Grade Recommendation - Reduce number of assessments? Allow districts to use at no cost. Legislative constraints "require." SHOULD LEAs decide to use it.
- Testing prior to third grade versus not testing before third grade? Measuring prior to 3rd grade. Program evaluation? Portfolio student.
- Recommendation #9 –
 - Universal Screener - Formative assessment. LEA option? Not a required 2nd grade option.
 - We are moving toward #9 - the reports are moving in the right now
 - Reporting that students understand as well in #9.
- Recommendation #10 - "Other methods" career inventories. Rambling and needs to be slimmed down. Need to highlight to elimination and EXPLORE/PLAN.
- Recommendation #11 - One ACT retake option? ED versus? Everyone? Below 21? Legislative options. Require action on the part of the student for the second take.
- Assessment? Module on assessment? For teacher prep programs - follow-up. Practicum & student teaching. Professional learning for current teachers?
- Further analysis: Duplicative? Local level deeper assessment? Which districts are doing what assessment? Are they aligned to standards? Is 11 hours for statewide annual assessment too much? Testing culture and test prep as a working hypothesis.
- Eliminating - too much assessment?
- Combine #14/#15 recommendations. Helping students and teachers.
- Recommendations - Grouping/organization. Themes?
- Next steps? What happens? Which will get implemented? Who will take on this work? Next steps owners/groups?
- Other groups: Scheduling & Logistics, Parent Group, & Higher Education workgroup?

- **Vicki's group feedback on the ATF report:**
 - Generally, the group I facilitated felt very positive about the report. Comments included:
 - It stated definitively what the state values and provides a clear “why” from the state’s perspective.
 - This is conscientious, and it gives us a guide. (Also appreciated the open nature of the task force).
 - Liked the specific nature of the recommendations; no jargon; not sweeping generalizations
 - Misspell on page 5 – Should be Nancy Ash (not “Ashe”)
 - Page 12, Assessment Principles, Summative Assessments, #5: Group suggested inclusion of career language here as well
 - Recommendation 7:
 - Group had a discussion about student accountability at this point. Raised the question about using quick scores (or some other score from EOC) on students’ final grades; expressed concern that, currently, this score elevates the student’s grade; wondered if there is any research demonstrating that inclusion of the students’ scores on grades causes them to put forth more effort.
 - Recommendation 8:
 - Purpose of portfolios should be clearly explained. One member of the group wanted them to be used to track student growth; currently, plans are to use samples of portfolios in a class for teacher accountability. Also, clarity is needed about what portfolios are already in use and how these connect with 2nd grade assessment.
 - Recommendation 10:
 - Group had questions about the meaning of “other methods”; expressed concern that responsibility for this would be shifted to districts; need clear, specific language here.
 - Recommendation 11:
 - Also, with regard to student accountability, recommend that students must take some sort of action (online prep; ACT prep class; etc.) between first ACT and retake in order to qualify for cost of 2nd administration.
- **Stephen's group feedback on the ATF report:**
 - Recommend inserting an Executive Summary
 - Whether in the State-Required subsection or somewhere else, recommend inserting the actual student hours spent taking the assessment. Think it is important to distinguish clearly what is spent on assessment and what is

spent by LEAs in prepping. We don't have that anywhere, but it is a very critical component

- Either after Recommendations or with Conclusions, the report needs a “Next Steps,” which will lay out tangible actions, which may include identifying specific parties that will be charged with assignments/action items. Right now, the report leaves the reader hanging.
- In the executive summary, note that the taskforce is recommending the dissolving of some existing assessments – thereby recommending the reduction of required/recommended state assessments
- Need a subsection title to draw attention to the creation of a task force and its charge. It is lost in the current narrative format.
- “Task Force Organization” (page 6) - Different word here? Organization seems incorrect usage. Group did not have another suggestion – ran out of time.
- Recommendations:
 - Suggest “grouping” the recommendations by likeness or another type of theming. They seem “scattered around” currently
 - Whether tied to a recommendation or placed somewhere else, there needs to be clear expectations that students will be shown/will see their test results (the new individual/and composite versions). Fear is still that the end consumer/the student will still be kept in the dark as to how he/she is performing.
 - Group wondered if this would be the place to identify which organizations would/should take each recommendation on – who would be responsible for tackling each recommendation
- Recommendation 11:
 - Group likes the intent; however, it needs more fleshing out
- Recommendation 12: Need to make sure this is for tested/ non-tested teachers
- Recommendation 14 and 15:
 - Also, language around the “transitioning away” was raised. This could be better addressed through the combining of the two.
- **Communications Plan, Ashley Ball**
 - Feature members of the group on social media (Twitter and Facebook)
 - Tentative planning to release the report on September 3 (Kickoff of the fall Classroom Chronicles kick-off)
 - Group gets to see one more preview before it is released
 - Parent Guide – www.tn.gov/tnready
 - Task force closed with Stephen Smith thanking the task force and asking them to complete an Assessment Task Force survey.

Appendix 3

Other Resources



Assessment Practices Task Force

Homework:

Between meeting 1 and 2:

- 1) Please Read "A Framework for Considering Interim Assessments" located here http://www.nciea.org/publications/ConsideringInterimAssess_MAP07.pdf.
- 2) Please choose one of the following stakeholder groups:
 - Students
 - Parents
 - Teachers

Talk to at least three people from this stakeholder group to better understand their perspective on the following questions:

1. What are your perceptions of assessments?
2. Do you think (you / your student(s)) takes the right number of tests? What would be about right?
3. What do you wish assessments could tell you? What would make them better?

A Framework for Considering Interim Assessments

Marianne Perie

Scott Marion

Brian Gong

National Center for the Improvement of Educational Assessment

February 13, 2007

Previous versions of this paper were presented at the 2006 Reidy Interactive Lecture Series (RILS) and to various CCSSO-sponsored State Collaborative on Assessment and Student Standards (SCASS) groups.

A Framework for Considering Interim Assessments

Marianne Perie, Scott Marion, and Brian Gong

The standards-based reform movement, first encoded in federal law as a result of the Improving America's Schools Act of 1994 (IASA), has resulted in the wide-spread use of summative assessments designed to measure students' performance at specific points in time. Under IASA, testing was required at three grades: once each at the elementary, middle, and high school levels. The enactment of the No Child Left Behind Act (NCLB) of 2001 required a significant increase in the prevalence of these large-scale summative tests. Policymakers' goal for most of these assessments is to measure students' knowledge and skills against some level of desired performance, such as attaining the level of *Proficient* or *Distinguished* or simply *meeting the standard*. While many have hoped that these end-of-year tests would provide instructionally useful information for educators, educators and others know this is not occurring. This is not because there is something "wrong" with these summative accountability tests, rather that they were not designed to meet instructional purposes. For example, these tests—by design—usually are administered as late in the year as possible and the results, by no fault of the assessment vendors, are returned after the students are home for the summer. In addition, the reports often provide only total score and performance level information for each student. Therefore, educators and policymakers have realized that other forms of assessments are necessary to provide information to inform instruction during the school year. Educators want to measure student progress toward important end-of-year goals and to receive sufficient information to determine what steps can be taken to further students' learning and achieve these goals.

This need for measuring student performance throughout the year has resulted in a rapidly growing influx of products in the field. Large numbers of vendors are selling assessments to states and districts that they call "benchmark," "diagnostic," "formative," and/or "predictive" with promises of improving student performance and helping schools meet their goals of showing adequate yearly progress or increasing pass rates on high school exit exams. A good district-level assessment can be an integral part of a state's comprehensive assessment system, used in conjunction with classroom formative assessments and summative end-of-year assessments. Yet, there is little research that these commercially-available assessments positively affect student achievement. In fact, many of these products cite the research on classroom formative assessment to indicate that their assessments improve student learning. However, few, if any, of these commercial products are the types of products or activities described in the Black and William (1998) analysis—the research most commonly cited. There is a growing concern among researchers that states and districts are buying assessment systems that promise to provide information to improve learning without fully examining the validity of these claims.

Policymakers and educators using assessments need to understand the purpose and limitations of any assessment. Is the purpose of the assessment to predict how students are likely to perform on an end-of-year assessment, to diagnose gaps in learning, to indicate the extent of student mastery of specific content and skills, or to evaluate a particular program or pedagogy? Because these assessments cost money and instructional time, they must provide experiences and information that are not available on the state large-scale assessment. At the same time, they should provide information that can be aggregated across students, occasions, or concepts

to provide information to those outside of the classroom while still aligned with information gathered through formative assessments within the classroom. Assessments that fill the gap between classroom formative assessments and state summative assessments are an integral part of any comprehensive assessment system and should be evaluated as such.

The purpose of this paper is to provide a framework for evaluating these mid-level assessments, which we call interim assessments, to help state and district leaders thoughtfully examine the commercially-available products or develop strong specifications for a customized system. This is a very large field and we are focusing primarily on those products that are currently being marketed for use by schools and districts, but we intend for the guidelines provided here to be useful for states or districts who want to develop their own interim assessments.

We began this work with discussions with state assessment leaders, assessment researchers, and others and found there was a consistent call for definitions of these different types of assessments—formative, interim, etc. Definitions, by their very nature, encourage analysis of each word and phrase within the definition, so creating a definition for interim assessment was not an easy task, particularly because we believe the definition for interim assessments should be driven by their purpose. We offer a definition in the next section, but throughout the paper our primary focus is on how these assessments are used. As such, we argue that the actual definition is always contextualized within specific purposes and uses. Consider, for example, how an assessment can be used summatively (to evaluate learning at the end of a unit), formatively (to inform subsequent instruction), or predictively (as an early warning system for future performance), to name a few purposes. We believe the purpose should be the main driver in defining and evaluating an assessment. Our discussion will focus on how the interim assessment fits into the comprehensive system and what unique purpose(s) the interim assessment serves. Therefore, after attempting to define these assessments and provide a classification system to consider their uses, we examine the characteristics of effective interim assessments, discuss the different purposes these assessments may serve, provide information on how to choose the best type of assessment for a given situation, and then offer guidance on evaluating the products that already exist in the marketplace.

Finally, we believe that there can be a difference between formative assessment and assessments used for formative purposes. Although we are strong proponents of the use of good formative assessment strategies, we also argue that interim assessments can serve formative purposes without meeting all the requirements for formative assessment. Furthermore, we recognize that not all interim assessments are designed to serve formative purposes and that other purposes may be legitimate, depending on the user's needs. Our goal is to provide a framework to help state and district-level consumers critically evaluate these systems.

Distinguishing Among Assessment Types

Before we can begin a thoughtful discussion on interim assessments, we need to agree on some definitions for the various names and types of assessments being used with the promise of improving student learning. The simple definition most first-year graduate education courses teach is that there are two types of assessments: summative, which are given at the end of instruction to provide information on what was learned, and formative, which are given at the beginning or in the middle of instruction to provide information about what the student knows

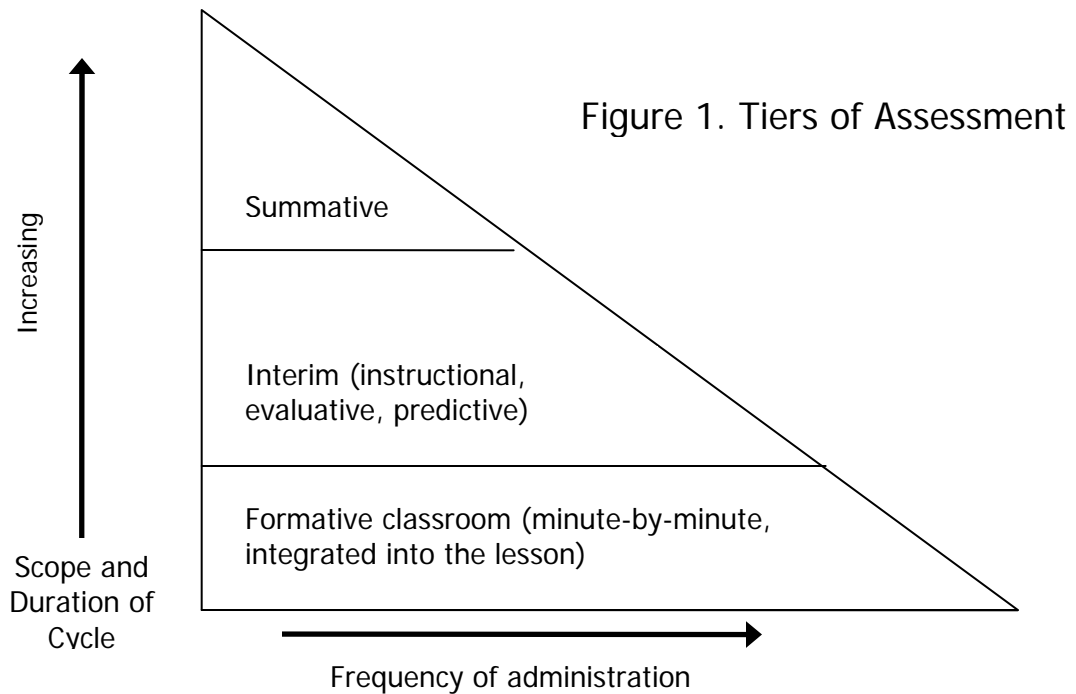
or doesn't know relative to what s/he should know at that point. However, this distinction of naming an assessment based on whether it's given at the beginning, in the middle, or at the end of instruction has led to many assessments being called formative even when they serve purposes that have little to do with providing useful information to teachers or students on improving student learning. In actuality, this distinction of formative and summative was initially used in the field of program evaluation with the express purpose of formative evaluation to provide mid-cycle corrections while summative evaluation was used to determine whether a program's results matched its goals (Scriven, 1967).

Our schema places assessments into three categories—summative, interim, and formative—and distinguishes among the three types based on the intended purposes, audience, and use for the information, not simply as a result of when the assessment is given. Summative assessments are given one time at the end of the semester or school year to evaluate students' performance against a defined set of content standards. These assessments are usually given statewide (but can be national or district) and are often used as part of an accountability program or to otherwise inform policy. They are the least flexible of the assessments.

Interim assessments may be administered on a smaller scale, typically school- or district-wide. While the results may be used at the teacher or student level, the information is designed to be aggregated at a level beyond the classroom level, such as to the school or district level. That is, they may be given at the classroom level to provide information for the teacher, but a crucial distinction is that these results can be meaningfully aggregated and reported at a broader level. These assessments may serve a variety of purposes, including predicting a student's ability to succeed on a large-scale summative assessment, evaluating a particular educational program or pedagogy, or diagnosing gaps in a student's learning. It is these purposes that determine the necessary features of the assessments.

The final type, formative assessment, is one given in the classroom by the teacher for the explicit purpose of diagnosing where students are in their learning, where gaps in knowledge and understanding exist, and how to help teachers and students improve student learning. The assessment is embedded within the learning activity and linked directly to the current unit of instruction. It can be a five-second assessment and is often called "minute-by-minute" assessment or formative instruction. Providing corrective feedback, modifying instruction to improve the student's understanding, or indicating areas of further instruction are essential aspects of a classroom formative assessment. There is little interest or sense in trying to aggregate formative assessment information beyond the specific classroom.

These three tiers of assessment—summative, interim, and formative—are shown in Figure 1. The triangle illustrates that formative assessments are used most frequently and have the smallest scope (i.e., the narrowest curricular focus) and the shortest cycle (i.e., the shortest time frame, typically defined as 5 seconds to one hour), while summative assessments are administered most frequently and have the largest scope and cycle. Interim assessments fall between these other two types on all dimensions.



Defining Formative and Interim Assessments

Because many of the commercial products available are marketed as “formative assessment” we feel it is important to go further in the definition and clearly distinguish between our use of the terms formative assessment and interim assessment. We understand that there is no current consensus on the definition of formative assessment, although leaders in our field have been working on a definition for some time now. Even less time and effort have been spent defining interim assessments. We offer a definition of interim assessment on the next page, but we fully expect it to be challenged and revised over the next several months. However, we felt it was important to lay out our current thinking on the definitions for both types of assessments before exploring the characteristics and evaluative criteria for interim assessments.

The definition of formative assessment, as proposed by a group of educational researchers wrestling with this issue¹ is as follows:

An assessment is formative to the extent that information from the assessment is used, during the instructional segment in which the assessment occurred, to adjust instruction with the intent of better meeting the needs of the students assessed.

¹ The group of educational researchers includes Jim Popham, Dylan William, Lorrie Shepard, Rick Stiggins, Scott Marion, Phoebe Winter, Don Long, Stuart Kahl, and Brian Gong.

It was further revised by state assessment and other education leaders attending the Formative Assessment for Students and Teachers State Collaborative in Assessment and Student Standards (FAST SCASS) meeting in Austin, Texas in October 2006 to read:

Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievement of intended instructional outcomes.

Although other definitions exist, both of these definitions fit nicely with the work done by Black and Wiliam who defined formative assessment as just one part of formative instruction. In their seminal piece, *Inside the Black Box*, Black and Wiliam (1998) argue that formative assessment cannot stand alone but must be a part of a whole system that uses the information from the assessment to adapt teaching to meet the learner's needs. In this definition of formative assessment, the assessment system consists of three phases:

1. Assessment (item development and delivery)
2. Diagnosis (analysis and reporting)
3. Prescription (pedagogy and professional development)

That is, a true formative assessment system does not stop with the development and administration of a test, but includes analyses that probe more deeply into what an incorrect answer implies about student learning and what should be done next or in the near future to further that learning. As stated by Black, et al. (2002), the first priority of a formative assessment is "to serve the purpose of promoting pupils' learning."

Both the classroom assessments, as described by Black and Wiliam and some commercially-available assessment systems may fit this definition of formative. However, we will distinguish between these small-scale, short-cycle assessments described most recently by Wiliam (2006) and the medium-scale, medium-cycle assessments currently in the field, which we are calling *interim assessments*. Although some have commented that since there are so many names for these assessments already in the field, the last thing we need is another name, we propose the term "interim assessment" as an umbrella term for all of these other assessments. We chose this term because one of the dictionary definitions of "interim" is, simply, "at an interval." Thus, benchmark, diagnostic, predictive, and even some formative assessments are considered interim assessments under our definition as follows:

Assessments administered during instruction to evaluate students' knowledge and skills relative to a specific set of academic goals in order to inform policymaker or educator decisions at the classroom, school, or district level. The specific interim assessment designs are driven by the purpose and intended uses, but the results of any interim assessment must be reported in a manner allowing aggregation across students, occasions, or concepts.

Typically, interim assessments are given several times a year, although a test that was administered once at some midpoint during the year could also be considered interim. By this definition, end-of-chapter tests available in most textbooks could be considered interim, if they are used in the aggregate. Teacher-created tests given at the end of a unit could be interim or formative, again depending on their purpose and design. The key components of the definition are that interim assessments (1) evaluate students'

knowledge and skills relative to a specific set of academic goals, typically within a limited time frame, and (2) are designed to inform decisions at both the classroom and beyond the classroom level, such as the school or district level. We will argue that the academic goals should be linked to the curriculum that a student has been taught (or should have been taught) at the time the assessment has been given.² Typically, these goals are aligned with state content standards. In addition, a key distinction among the assessment types is the recipient of the information. For interim assessments, although the results may be used by the teacher to adjust instruction, another recipient could be an educational leader such as a school administrator, curriculum supervisor, or district policymaker. Unlike formative assessments, interim assessments can be aggregated easily, and the results can be examined across classrooms, schools, or even districts.

Overview of Interim Assessments

There are many forms of interim assessments currently available, often labeled “benchmark,” “formative,” “diagnostic,” or “predictive.” They can be given early in the school year, mid-way through, or periodically throughout the year. The one common thread is that they are designed to give information about the students’ level of knowledge and skills before the end of the school year. Our goal is not to describe one assessment, called an interim assessment, but to focus the discussion on the different uses of these assessments and describe how the best practices for any interim assessment are related directly to the intended use of the assessment.

We encourage the reader to think broadly about the possible forms of interim assessments, from commercially-purchased, computer-based sets of multiple-choice items to more locally created sets of extended performance tasks administered somewhat commonly throughout a school, district, or state. We do not intend to tout one type of interim assessment as being the best—although we argue that some are clearly superior for improving learning than others—but to encourage users to be explicit about the desired purpose of the assessment and then find the assessment that best fits that purpose. For example, an interim assessment may be given in order to

- ✓ Evaluate how well the student has learned the material taught to date,
- ✓ Predict students’ performance on a summative assessment, or
- ✓ Determine whether one pedagogical approach is more effective in teaching the material than another.

These are just three possible purposes. Interim assessments may serve multiple purposes as well, by providing aggregate information on student achievement at a district level, while providing specific feedback on where the gaps in a particular student’s knowledge are at the classroom level. Many currently-available interim assessments have been called “early-warning tests” or, more pejoratively, “mini-summative tests.” Their purpose is to determine whether the student is on track to succeed on the summative assessment. These tests may also serve formative purposes as they should diagnose and provide corrective feedback to help the student get on track to succeed on the summative assessment and not to simply predict how

² One exception to this rule of assessing what has already been taught is if the purpose of the assessment is to determine the starting point for instruction based on the knowledge and skills students already have when the students’ level is unknown to the teacher or school.

the student will perform on the end of year test. Interim assessments may serve other purposes as well, including motivating and giving feedback to students about their learning. Many students think they know something but are often surprised and motivated to do something different when they get the results back from a quiz or classroom test, even without specific feedback.³ Another purpose of an interim assessment could be to provide information to help the instructor better teach the next group of students, by evaluating the instruction, curriculum, and pedagogy.

Exhibit 1 provides a more comprehensive list of possible purposes for interim assessments, as well as information on the types of information that may be assessed and at what level the information will be used. In addition to the purposes discussed in the previous paragraphs, we also considered other purposes, such as ensuring that teachers are staying on track in terms of teaching the curriculum in a timely manner, providing a more thorough analysis of the depth of students' understanding, and determining whether students are prepared to move on to the next instructional unit, to name a few.

Summarizing this large table brought us to three general classes of purposes for interim assessments: Instructional, evaluative, and predictive. Although this categorization is not perfect, it seems to capture the essence of most of the goals of using an interim assessment system. Further, we recognize that many assessments are not designed to serve only a single purpose, but we argue that few assessments or assessment systems can serve more than two or three purposes well and they tend to work best when the various purposes have been prioritized explicitly. Thus, an important additional step is to check not only whether the assessment is being used for its intended purposes, but to check the quality with which it meets its purpose.

³ Although the research (e.g., Sadler, 1989) is quite clear that task-specific feedback is superior for improving student performance.

Exhibit 1: Interim Assessment Chart

Purposes: How results are used	What is measured	Who the results are primarily intended to inform					When the test is given			Type of item that would be appropriate						
		State policy makers	District policy makers	School Administrators	Teachers	Students & parents	Early in the instructional period	Periodically- on a calendar-based interval (e.g., every X weeks)	Periodically - at an instructional break (e.g., end of a unit)	MC	SCR	ECR -- essay	Short tasks	Synthesis task (research report; lab design/ analysis)	Checklist	Oral explanation
Instructional																
To provide aggregate information on how students in a school or district are doing at a given point in the school year	Content standards of material that is supposed to be covered to date	✓	✓	✓	✓	✓		✓			✓	✓	✓	✓		
To provide feedback on individual students for teachers and aggregate data on student progress for schools and districts	Knowledge and skills of material to be covered next		✓	✓	✓				✓		✓	✓	✓	✓		

	Who the results are primarily intended to inform	Who the results are primarily intended to inform			When the test is given			Type of item that would be appropriate								
		State policy makers	District policy makers	School Administrators	Teachers	Students & parents	Early in the instructional period	Periodically - on a calendar-based interval (e.g., every X weeks)	Periodically- at an instructional break (e.g., end of a unit)	MC	SCR	ECR -- essay	Short tasks	Synthesis task (research report; lab design/ analysis)	Checklist	Oral explanation
Purposes: How results are used	What is measured															
To diagnose gaps between student knowledge and intended curriculum	Information that has been presented in the classroom to date.			✓	✓	✓	✓		✓	✓	✓	✓				✓
To see where a student's current knowledge and skills are in relation to what is about to be taught. i.e., is the student ready to learn the material in the next instructional unit?	Knowledge and skills required to understand upcoming curriculum as well as samples of upcoming curriculum			✓	✓	✓	✓			✓	✓	✓	✓			✓

Purposes: How results are used	What is measured	Who the results are primarily intended to inform					When the test is given			Type of item that would be appropriate						
		State policy makers	District policy makers	School Administrators	Teachers	Students & parents	Early in the instructional period	Periodically-on a calendar-based interval (e.g., every X weeks)	Periodically - at an instructional break (e.g., end of a unit)	MC	SCR	ECR -- essay	Short tasks	Synthesis task (research report; lab design/ analysis)	Checklist	Oral explanation
To determine students' ability levels in order to group them for instructional purposes	Knowledge and skills required to understand upcoming curriculum as well as samples of upcoming curriculum			✓	✓	✓	✓			✓	✓		✓			✓
To help the teacher determine where to put his/her efforts: reviewing previously taught material, teaching material in more depth, moving on to new unit.	Knowledge and skills of material just taught			✓	✓	✓					✓	✓	✓			✓

Purposes: How results are used	What is measured	Who the results are primarily intended to inform					When the test is given			Type of item that would be appropriate						
		State policy makers	District policy makers	School Administrators	Teachers	Students & parents	Early in the instructional period	Periodically - on a calendar-based interval (e.g., every X weeks)	Periodically- at an instructional break (e.g., end of a unit)	MC	SCR	ECR -- essay	Short tasks	Synthesis task (research report; lab design/ analysis)	Checklist	Oral explanation
To enrich curriculum	Information that has been presented in the classroom, but assessed in a way that requires students to show their depth of understanding			✓	✓	✓			✓			✓	✓	✓		✓
Test knowledge and skills not easily assessed in a time-constrained large-scale assessment	Information that has been presented in the classroom, but assessed using lab analyses or research reports or other such mechanisms				✓	✓			✓					✓		
To allow students to evaluate their knowledge and see areas in which they need to grow.	In depth knowledge and skills of material just taught				✓	✓			✓	✓	✓	✓	✓	✓		✓

Purposes: How results are used	What is measured	Who the results are primarily intended to inform					When the test is given			Type of item that would be appropriate						
		State policy makers	District policy makers	School Administrators	Teachers	Students & parents	Early in the instructional period	Periodically- on a calendar-based interval (e.g., every X weeks)	Periodically- at an instructional break (e.g., end of a unit)	MC	SCR	ECR -- essay	Short tasks	Synthesis task (research report; lab design/ analysis)	Checklist	Oral explanation
To encourage students to evaluate their own knowledge and discover the areas in which they need to learn more.	Knowledge and skills of material to be covered next				✓	✓	✓		✓	✓	✓	✓				✓
To practice for summative test	A portion of the knowledge and skills that will be measured on the summative assessment			✓	✓	✓		✓	✓	✓						
Evaluative																
To evaluate the effectiveness of various curricular and/or instructional practices	Content standards of material that is supposed to be covered to date		✓	✓	✓				✓							

Purposes: How results are used	What is measured	Who the results are primarily intended to inform					When the test is given			Type of item that would be appropriate						
		State policy makers	District policy makers	School Administrators	Teachers	Students & parents	Early in the instructional period	Periodically- on a calendar-based interval (e.g., every X weeks)	Periodically- at an instructional break (e.g., end of a unit)	MC	SCR	ECR -- essay	Short tasks	Synthesis task (research report; lab design/ analysis)	Checklist	Oral explanation
To increase teacher knowledge of assessment, content domain, and student learning	Content parallel to what is on summative assessment, although covering only what has been taught to date			✓	✓		✓		✓	✓	✓	✓				
To reinforce curricular pacing	Content standards of material that is supposed to be covered to date		✓	✓	✓	✓		✓	✓	✓	✓	✓		✓		
To provide information on how best to target the curriculum to meet student needs	Knowledge and skills required to understand upcoming curriculum as well as samples of upcoming curriculum		✓	✓	✓		✓	✓	✓	✓	✓	✓				✓

Purposes: How results are used	What is measured	Who the results are primarily intended to inform					When the test is given			Type of item that would be appropriate						
		State policy makers	District policy makers	School Administrators	Teachers	Students & parents	Early in the instructional period	Periodically- on a calendar-based interval (e.g., every X weeks)	Periodically- at an instructional break (e.g., end of a unit)	MC	SCR	ECR -- essay	Short tasks	Synthesis task (research report; lab design/ analysis)	Checklist	Oral explanation
To predict student achievement on summative test	Same content standards as on summative tests, as mini versions of the summative assessment	✓	✓	✓	✓	✓		✓		✓	✓	✓				
To predict student achievement on summative test	Same content standards as on summative tests, assessing appropriate segments that should have been taught to date	✓	✓	✓	✓	✓		✓		✓	✓	✓				
To provide an early warning system for students who are not on track to succeed on X	Knowledge and skills required for success on X (e.g., assessment anchors on summative assessment)	✓	✓	✓	✓	✓		✓		✓	✓					

Instructional

The primary goal of an interim assessment designed to serve instructional purposes is to adapt instruction and curriculum to better meet student needs. Of the three purposes, this one aligns most closely with the previous definitions of formative assessment. That is, the results of these assessments are used to adjust instruction with the intent of better meeting the needs of the students assessed. However, the testing and reporting time frame of these interim assessments is typically medium-cycle, whereas classroom formative assessments tend to operate on shorter cycles.

Subsumed under this purpose are other types of assessment that certainly would not meet the definition of *formative* presented earlier, but are instructional nonetheless. Consider, for example, an assessment that asks a student to explore a concept in greater depth or one that provides tasks that stretch students and teachers to do things at deeper cognitive levels than they might otherwise. These purposes would certainly be considered instructional, and we would argue that these are laudable goals. However, this type of assessment does not neatly fit the earlier definition of a formative assessment that *"information from the assessment is used, during the instructional segment in which the assessment occurred, to adjust instruction with the intent of better meeting the needs of the students assessed"* or that it *"provides feedback to adjust ongoing teaching and learning to improve students' achievement of intended instructional outcomes."* Rather, the assessment itself provides the rich instruction. It is worth noting, however, this type of assessment meets an earlier definition of formative provided by Black, et al. (2002) stating that the first priority of a formative assessment is "to serve the purpose of promoting pupils' learning."

As a second example, consider the features included in many commercially-available systems. A typical system contains a bank of items nominally aligned with the state curriculum that teachers can use to create a test to evaluate student learning on the concepts taught to date. Results are reported immediately, and data are disaggregated by content standard allowing teachers to identify strengths and weaknesses in the students' learning. This type of interim assessment might be labeled formative, but we would argue that to be truly formative it must be fully aligned with state curriculum and provide more in-depth analyses of student misconceptions or lack of understanding along with instructional tools and strategies for improving instruction. However, this type of interim assessment does fall under the instructional category. Also, because the results can be aggregated and used at a level outside of the classroom, it meets the definition for interim.

As a third example, consider an end-of-chapter test that students self-administer when they have completed a unit. The purpose of this assessment may be two-fold—to ask students to assess their understanding of a particular topic and to provide information that can be aggregated at the classroom level to inform the teacher of the readiness of the class to proceed to the next unit.⁴ This type of assessment is certainly an interim assessment serving an instructional purpose, although we would argue that there needs to be evidence that this end-

⁴ The results of this assessment could also be aggregated across any classroom or school that uses the same textbook.

of-chapter test is fully aligned with state standards if the test is also to serve the purpose of preparing students for the end-of-year assessment.

We also argue that to serve instructional purposes an assessment system must go beyond simply providing data. Test designers must provide strategies for interpreting and using the data to effectively modify classroom instruction. At the very least, the assessment system should include information about possible interpretations for each incorrect answer. It is worth noting a tension between the need for professional development to accompany these assessment systems and the ownership of that responsibility. It is the contention of many assessment developers that tools and strategies for improving instruction are the teacher's responsibility, not the instrument provider's. Many policymakers, however, want to see that professional-development support included in the package that they purchase. We lean toward the latter viewpoint in that an assessment system purchased for instructional purposes must include professional development to ensure that educators have the tools to use the results appropriately. We believe that this should be a shared responsibility among the developer and the user.

Evaluative

Another type of purpose an interim assessment might serve is to provide evaluative information about the curricular approach or instructional strategies. Think of this as a programmatic assessment designed to change instruction not necessarily in mid-term but over the years. The students benefiting from the information gleaned from these assessments would not necessarily be the students assessed, but the students that would receive the instruction in the future.

Assessments used for evaluative purposes could be given district wide to compare the effectiveness of various instructional programs for improving student learning. Consider, for example, a district that is experimenting with more than one reform program or pedagogical strategy across different schools. The use of interim assessments in this context could be an effective way of monitoring the relative efficacy of each program. Likewise, assessments could be given at various points throughout the year to measure growth—not with the intention of intervening but for evaluating the effectiveness of a program, strategy, or teacher.

The assessments could also be used on a smaller scale, providing information on which concepts students understood well and which were less clear to teachers within one or more schools with the goal of helping them modify the curriculum and instructional strategies for future years. Other purposes could be to provide a more in-depth understanding at the school level on how the test items link to the content standards and how instruction can be better aligned with improved performance on the test. Of course, teachers can and should always learn from their experience. Any instructional interventions that could improve instruction in a current year should be implemented.

In our definition, an *evaluative* assessment would be designed explicitly to provide information to help the teacher, school administrator, curriculum supervisor, or district policymaker learn about curricular or instructional choices and take specific action to improve the program, affecting subsequent teaching and thereby, presumably, improving the learning. Assessment systems designed to serve evaluative purposes must provide detailed information about relatively fine-grained curricular units. A global, content area score will not suffice. However, not every student needs to be assessed in order for the teacher or administrator to receive

high-quality information from the assessment. A matrix sample could be used to maximize the information while minimizing the time spent on assessments in the classroom.

Predictive

We suspect that there are few assessment systems where the only purpose for the system is to predict performance on some later assessment. Nevertheless, the predictive purposes of interim assessments are important to many users and this interest could increase as the annual NCLB targets continue to rise. In addition, assessments in this category could be used to predict performance on the high school exit exam or success with post-secondary curriculum. Although our focus will be on the predictive use of these assessments, we expect most users want additional information to help them improve the performance of students for whom failure is predicted.

These predictive assessments are designed to determine each student's likelihood of meeting some criterion score on the end-of-year tests. End users should be able to aggregate the results to the classroom, subgroup, school, and even district level.

Although there has been some discussion as to the worth of an assessment that only provides information that a student is on track to fail without additional diagnostic information to guide interventions, we have received anecdotal reports from certain users that scores from these predictive-type assessments can serve as a screener. That is, in some districts predictive tests are used solely to identify students who are not on track to succeed on the end-of-year assessment. Then, once those students are identified, they are given further probes to determine areas of weakness and provided with remedial instruction, extra support, and/or tutoring. This scenario could be an example of how interim and formative assessments work together to help improve student performance on a summative assessments. It also highlights the importance of having all three of these assessment types aligned in a comprehensive assessment system.

A confounding variable on any predictive test is that if it provides good feedback on how to improve a student's learning, then its predictive ability is likely to decrease. That is, if the test predicts that a student is on-track to perform at the basic level, and then appropriate interventions are used to bring the student to proficient, the statistical analysis of the test's predictive validity should underpredict student performance over time. However, it is important to track the performance of students predicted to succeed on the summative test. That is, it should not be considered a strike against the predictive test if a student predicted to fail the summative test actually passes it, but questions should be raised if too many students predicted to pass the summative test actually fail it.

Identifying the Goal

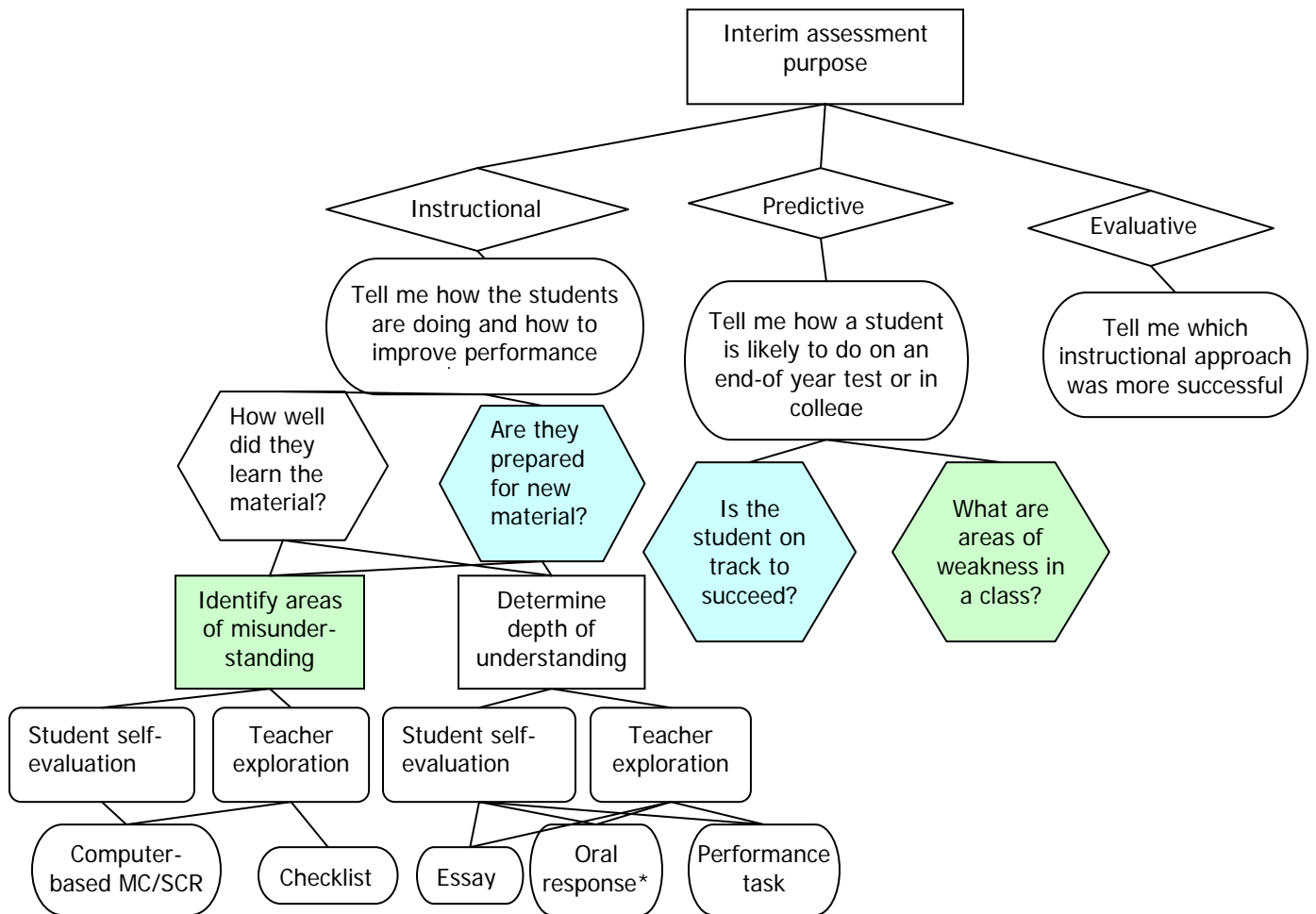
As policymakers decide to bring an interim assessment system to their state/district/school we encourage them to have a theory of action for how the particular assessment system will work in the teaching-learning cycle. As a start, we think it will be helpful for educational leaders to address the following questions:

1. What do I want to learn from this assessment?
2. Who will use the information gathered from this assessment?
3. What action steps will be taken as a result of this assessment?

4. What professional development or support structures should be in place to ensure the action steps are taken?
5. How will student learning improve as a result of using this interim assessment system and will it improve more than if the assessment system was not used?

The answers to these questions will dictate the type of assessment needed and will drive many of the design decisions including the types of items used, the mechanism for implementing it, the frequency with which it should be administered, and the types of reports that will need to be developed from the data. We present a partial decision tree as shown in Figure 2, which shows the types of questions policymakers should be able to answer about the type of system they want before developing or purchasing one. Importantly, these questions and the associated answers serve as the beginning of a validity argument in support of (or to refute) the particular assessment system. Note that it is not complete, as a complete tree would need several pages for the display, but it should provide a clear example for policymakers to use when thinking through the issue of selecting or designing an interim assessment system.

Figure 2: Decision Tree to Use in Creating an Interim Assessment System



*May be better answered using classroom formative assessment techniques.

Answering the questions in this decision tree also may reinforce the idea that it is often appropriate to consider multiple purposes in designing or choosing an interim assessment system. For instance, while the primary purpose of giving an interim assessment may be evaluative, we would hope that given the results for a specific set of current students, teachers and school leaders would attempt to provide remediation programs for students not understanding key concepts. Likewise, even when the primary purpose of an interim assessment is to predict success on the end-of-year assessment, a policymaker may also want the predictive assessment to provide some diagnostic information so that educators can intervene with students predicted to score below a critical level. Areas of overlap in the questions are indicated by the coloring schemes. Again, some of the instructional questions may be answered best through the use of classroom formative assessments, but there are interim assessments that may also provide useful information at the classroom level and also answer predictive or evaluative questions.

The next section describes the various characteristics of any effective assessment system. It also provides several possible approaches to developing one. The appropriate approach should be determined based on the answers to the above questions. There is no one-size-fits-all assessment, only a best design for a desired use and the existing constraints and resources. We believe that many educational leaders consider a cost-benefit relationship before investing in such a system, but we fear that the equation often tips in favor of the costs. For instance, it is cheaper to score multiple-choice items than constructed-response items or performance tasks, and it often costs less to buy a computer-based testing system than to invest in professional development for all teachers. We recognize the reality of constrained budgets, but argue that saving a few dollars on an assessment system might actually “cost” more in terms of opportunities for learning that may be lost as a result of cutting up-front purchase costs.

Characteristics of an Effective Interim Assessment System

This section is intended to help educational leaders either choose or develop a strong interim assessment system for their schools. We recognize that some districts or states will be looking to purchase an already-available assessment system, while others will be looking to create a system customized to their needs. The considerations described below are appropriate both for evaluating currently-available systems and for designing new systems.

Shepard's (2006) requirements for an assessment to have strong formative potential include:

- ✓ Can never be *all* multiple-choice.
- ✓ Must provide *qualitative* insights about understandings and misconceptions not just a numeric score.
- ✓ Should have immediate implications for what to do besides re-teaching every missed item (the 1000 mini-lessons problem).

While we continue to separate formative assessment from interim assessments with formative purposes, we believe that any good interim assessment system will produce results that can be used to inform instruction. We are clear that the current research literature supports the types of formative assessments defined earlier and does not yet offer guidance about the efficacy of certain interim assessments. However, certain reports of best practices (e.g., Marshall, 2006) indicate that effectively using interim assessments can be a lever for powerful educational reform. Characteristics of an effective interim assessment system include:

- ✓ A rich representation of the content standards students are expected to master
- ✓ High quality test items that are directly linked to the content standards and specific teaching units
- ✓ A good fit within the curriculum so that the test is an extension of the learning rather than a time-out from learning
- ✓ Reliable results that are easy to interpret and clear guidance on how to use the results

Other characteristics may be important depending on the purpose(s) of the assessment. Again, we emphasize that the purpose must be clearly stated before one can truly determine or evaluate the necessary characteristics of the assessment. One element we consider essential to any interim assessment system is validity evidence. The information provided by and uses of interim assessments should be validated, and we will focus on this aspect of the assessment system later in the paper.

One strategy for defining the desired characteristics is to focus on the reporting element. What do we want the tests to tell us? A report is a mechanism for translating the assessment data into action and should be one of the first considerations in designing a new assessment. In the next few pages we discuss important considerations for reporting results from interim assessments, and then we follow the assessment process from design to implementation and scoring.

Reporting Results

We would argue that the score report is one of the most important components of an interim assessment system. It serves to make the results actionable. A good report will indicate not only which questions a student answered incorrectly, but also what the student's incorrect response or set of responses implies about learning gaps or misconceptions, what further probes may be administered to gather more information about the problem, and instructional strategies for improving student learning on that topic. Again, this last point about instructional strategies may be provided in a supplemental professional development package, but we believe it should be considered when designing a reporting system. At the least, there should be research behind the items in the assessment that can be used to inform general instructional approaches that might be considered for improving student learning. That is, assumptions about where a student is in his/her learning progression based on an incorrect response should be grounded in research on learning progressions.

Student-Level Information

In evaluating an interim assessment program, it is important to examine the type of student-level information provided and potential users must evaluate the extent to which that information matches the goals of the program. Many commercially-available systems provide information on which items were answered correctly and incorrectly and how those items map to state content standards. Most systems also aggregate the items into clusters or subscores. That is, the results may be broken out by content strand, so a student may have an overall score for a math test and also have subscores in numbers and operations, geometry, and measurement, for example. Unfortunately, many assessment programs have had little success in validly reporting meaningful subscores, so we must be cautious with promises or assumptions that any interim assessment system can do so. Consider, however, the richness of student-level information that can be gained through performance tasks or essays that are reviewed by the

teachers, or even results of multiple-choice or short-answer tests when the questions have been written to measure specific features in a student's learning progression. These types of student-level information may be displayed qualitatively (in the form of notes) rather than quantitatively (number-correct scores) and could be quite valuable.

Aggregate Information

For this to be an interim assessment, it must be possible to aggregate information at the classroom, school, or district level. Given time constraints and limited resources, a teacher may want a more holistic view of the level of understanding of all the students in the classroom, rather than individual diagnostic information for each student. A district policymaker may be concerned about general trends by school or about the proportion of students expected to reach proficiency by the end of the year. It would be important in this situation to have a reporting system in place that could provide aggregate information across any level desired, from classrooms within schools to schools within districts to districts within a state. Following the example in the previous section, rich information about an individual student can be gained from an essay test, but the total score (or even the score on various components of a rubric) could be aggregated across students to provide additional information both at the classroom level and beyond.

Additionally, it may be important to be able to *disaggregate* the data across different groups. That is, it may be important to distinguish performance among different groups of students or across students in different schools or different instructional programs. One result of NCLB is that assessment results are almost always broken out by student groups such as race/ethnicity, gender, socioeconomic status, disability status, and English language proficiency. Many available interim assessment systems also include mechanisms for grouping results into these groups.

Follow-up Information

As discussed in an earlier section, those who wish to use an interim assessment to help inform instruction need to know where the students are in their learning, how that compares to where they are going, and how best to reach that goal. The reports must lend themselves to informing a plan for action intended to further the student's learning in appropriate areas. Simply aggregating the data and identifying the content strands on which the students missed the most items is not sufficient. Ideally the follow-up information will be curriculum-specific. For example, when interim assessments are not tied to specific curriculum sequencing, students might answer items incorrectly because (a) they were instructed on the content but did not understand a concept or (b) they had not yet covered the material at the point the test was administered. Given the money spent on some interim assessment packages, it does not make sense that a teacher's first interpretative action involves determining whether the missed items were taught or not.

Even if the purpose is predictive rather than instructional, the follow up information should include the expected end-of-year achievement for each student along with intervention strategies to improve students' end-of-year outcomes. That is, if a student is currently projected to score at the basic level, it would be helpful to know how to move the student towards scoring at the proficient level. This aspect of reporting is the most frequently overlooked but is arguably the most important. The recommendations included should be grounded in research on student learning and cognition and provide clear guidance to teachers.

Summary of Interim Report Features

Again, it is important to clarify the purpose of the assessment when examining its various features. We recommend visualizing and designing the intended reporting system as a way of clarifying all the information desired from the assessment. Assessments serving an instructional purpose will have different features in their reports than those serving predictive or evaluative purposes. Further, to judge a reporting system's effectiveness, it must be vetted through those who need to use the information: teachers in most cases but also school leaders.

Developing the Items

Once the purpose and desired form of results are clear, we can begin to think about the types of items that would be appropriate for the interim assessment. Different types of items may be selected for different purposes. For instance, if the test is being used to predict student performance on an end-of-year test, then the item types should match the types of items that will be found on that summative test—typically multiple-choice and short constructed-response items. If, however, teachers also want further information on a student's depth of knowledge or areas of weakness, other types of items may need to be used to supplement the results of the interim assessment, such as performance tasks or essays.

The items must be developed in such a way to represent in-depth coverage of the specific content standards and curricular goals to be assessed. The items should have all the positive elements of any item in a large-scale assessment (reliable, free from bias, etc.) but also be able to provide information on the student's depth of understanding or what a student does not know through an analysis of incorrect responses or a thorough scoring rubric. Items should be linked directly to both the state content standards and to the individual units of learning as determined by the curriculum guides.

Scoring is also an important consideration in item development. Many commercial systems are designed to be scored electronically. Certainly, electronic scoring allows results to be produced quickly and aggregated easily across classrooms and schools. However, one should consider the learning value of students self-scoring or teachers scoring student work. This is particularly true for open-ended items where examination of the raw student work may enable teachers to observe and interpret patterns in student responses that may be lost in a scoring guide. Scores can then be recorded and uploaded for aggregation across classrooms or schools. In the following sections, we discuss just a few types of items in terms of their appropriateness for an interim assessment as well as distinguishing characteristics of good items.

Multiple Choice

Shepard (2006) makes it clear that assessments used primarily to inform instruction should not consist solely of multiple-choice (MC) items because MC items do not provide enough information about *how* students understand. However, some multiple-choice items can provide constructive feedback on the breadth of students' knowledge as well as providing a quick check on misconceptions or incomplete understandings. Specifically, items built with an understanding of students' learning progressions will provide richer information on gaps in student knowledge. A classic example of a multiple-choice item that provides evidence of student learning is the following:

Consider the four diagrams shown below. In which of the following diagrams, is one quarter of the area shaded?

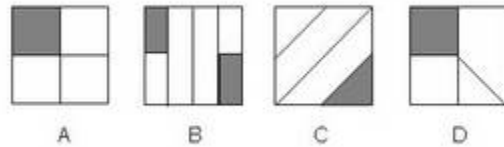


Diagram A is the obvious answer, but B is also correct. However, some students do not believe that one quarter of B is shaded because of a belief that the shaded parts have to be contiguous. Students who believe that one quarter of C is shaded have not understood that one region shaded out of four is not necessarily a quarter. Diagram D is perhaps the most interesting here. One quarter of this diagram is shaded, although the pieces are not all equal; students who rely too literally on the “equal areas” definition of fractions will say that D is not a correct response. By crafting questions that explicitly build in the under- and over-generalizations that we know students make, we can get far more useful information about what to do next.⁵

For items such as these to be instructionally useful, there should be a clear theoretical basis related to how students develop proficiency in the domain when developing the item distractors. There must be a sound research basis for linking each wrong answer to a specific gap in learning. In addition, as good as any one item may be, usually many multiple-choice items are required to gain real insight into why a student answers incorrectly.

Another way to enrich the information gained from multiple-choice items is to ask the student to justify their response to each item. Asking questions such as “why did you select that answer?” or “what rule or information did you use to arrive at your decision?” or simply asking students to explain their answer can provide additional insights into the student’s thinking. This strategy also allows the test to serve two purposes: to provide quick aggregate data on student performance and to provide more in-depth information about a student’s thinking process that a teacher can use to inform instruction.

Unfortunately, the type of multiple-choice item presented as an example here is rarely found in practice with most commercially available interim assessments. In fact, we would argue the same point regarding end-of-year large-scale assessments. While it is certainly possible to write high quality multiple-choice items, the demand for a large quantity of items produced quickly has led to considerable concerns with quality of multiple-choice items found on commercially-available assessments. Ideally, if the purpose of using an interim assessment requires large numbers of multiple-choice items for quick-turnaround and easily aggregated data, test developers should aim to include as many multiple-choice items as possible that are

⁵ This example is from a presentation by Dylan William delivered at ETS in 2006. The exact origin of the item was not given.

built from learning theory and provide useful information about wrong answers as well as correct ones.

Open Ended

Any type of item that requires students to generate their own answers should provide richer detail about student thinking as well as support the goal of moving students towards deeper thinking about the subject matter. Once again, however, to provide information that is useful to instruction, the items should be constructed in a manner that allows the student to provide information on his/her thinking. Some examples of these types of items include essay prompts that ask students to justify their point of view. An essay may take longer to score and even longer to extract all the information available to inform instruction, but it provides a rich diagnostic tool for teachers. Simply examining the student work as they score the essays can also provide teachers with professional learning opportunities by giving them greater insights into how their students process information.

If the purpose of the assessment is to provide aggregate information quickly, short constructed-response items may be more appropriate. However, in some cases more information can be obtained from a related set of items than from a single item. Consider, for example, a simple arithmetic problem. If a student answers a subtraction problem as follows

$$\begin{array}{r} 584 \\ -68 \\ \hline 524 \end{array}$$

all we know is that the student answered the item incorrectly. However, if we look at a *set* of student responses

584	495	311	768	821
-68	-73	-82	-34	-17
524	422	371	734	816

we now have more information to process. On the surface, we know that the student answered 2 of the 5 items correctly, but if we look closely, we see that the student makes the same error on the three items answered incorrectly. On items 1, 3, and 5, where the number on the bottom contains a digit that is higher than the one on the top, the student is simply reversing the order. That is, in the first item, the student should be subtracting the 8 from the 4 (carrying a 1 to make it 8 from 14), but instead, the student flips it so that s/he is subtracting the 4 from the 8. The same error is made on all three items, providing richer information on the type of corrective instruction needed.⁶

For these types of item sets to be effective, however, strong supporting material must be included as part of the assessment. That is, included in the assessment system should be instruction on the types of errors to look for, how to detect errors, and what corrective steps

⁶ An expanded version of this and other task sets can be found in Pellegrino, Chudowsky, and Glaser (eds.) (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press

can be taken once a problem is identified. Again, these common errors on student thinking need to be identified based on modern conceptions of learning and empirical research on the nature and distribution of such misconceptions.

Performance Tasks

Few, if any of the current commercially available interim assessment systems include items other than multiple-choice items. Part of the purpose of this paper is to broaden the discussion of interim assessments to include more locally-developed or other customized approaches. Adopting this viewpoint allows us to consider a wider range of item types than is typically the case with commercial systems. Performance tasks can be included as part of an interim assessment system to provide opportunities for more in-depth focus on the content area than is often the case with selected-response item types. These tasks can be relatively short, such as graphing the results shown in a table, or more complex, such as designing, carrying out, and analyzing a science experiment. Again, as long as the results can be aggregated and used at a level beyond the classroom, an assessment with these types of tasks falls under our definition of interim.

The tasks themselves can enrich and deepen the curriculum. That is, a task may require students to learn one particular objective in much greater detail by analyzing it or generalizing from it. These tasks can also serve the purpose of having students self-evaluate their learning and understanding of a concept or to synthesize their understanding across several concepts or standards. Reading about an idea in a book is very different than making it come to life in a task. Students will be able to better evaluate their own depth of knowledge and explore other avenues on their own, which serves a type of formative purpose.

Performance tasks, particularly extended tasks, can serve instructional purposes more readily than other interim assessment item types. Having students engage in these tasks provides opportunities for the teacher to observe students as they solve problems or otherwise work on the task. The teacher may stop and probe to understand better why the students are doing poorly or well. Depending on the larger purpose of the interim assessment system, this interaction may not be possible, but teachers can still conduct systematic observations to provide feedback to students soon after the task is completed as well as gather data that can be aggregated across students, concepts, or occurrences.

Another way to conceive of this type of performance item that can be easily aggregated in an interim assessment system is to consider a checklist. Under certain circumstances where a set of discrete tasks is well defined, a teacher may use a checklist to evaluate all students as they conduct these series of tasks. As the students work, the teacher may simply walk around the classroom with the checklist and mark off which students successfully complete each component of the task and note which aspects students are struggling to understand. The checklist can be aggregated across students or tasks and the notes can be used to improve instruction. In addition, the teacher may stop and probe to understand better why the students are doing poorly or well. We would encourage educational leaders to consider the types of tasks being assessed and determine whether a checklist or a rubric would provide more useful information.

Performance tasks, particularly extended tasks, such as a research paper, science laboratory, or historical debate, have the advantage of helping to erase the artificial boundaries between

assessment and instruction. When students are engaged in such tasks, an observer struggles to determine whether there is an assessment underway or simply an interesting instructional unit. Perhaps most importantly, these types of performance tasks, when well designed, increase student motivation by engaging them in meaningful interactions with rich subject matter.

Summary on Evaluating Item Types

Again, once the purpose and reporting elements are clearly outlined, the type of items required should become clearer. Assessments serving a predictive purpose might need to be more aligned with the test design of the summative assessments to which they are linked, while assessments serving an instructional or evaluative function will likely be better served by open-ended items and performance tasks. The turnaround time of results should also be considered. That is, if a teacher or educational leader needs results quickly, multiple-choice or short constructed-response items can be scored more rapidly than essays or performance tasks. The audience should also be considered. If a school or district administrator is only interested in obtaining aggregate information on the students, quick-turnaround responses, typically in the form of multiple-choice or short open-ended items, will suffice. If, however, an additional goal is to provide useful information back to the teachers to use in instruction, more performance tasks or longer open-ended items may serve both purposes.

Beyond the specific item types, users should carefully examine the quality of the items they are considering using. An assessment system will be unable to meet any of the goals put forth by stakeholders if the system is built on low-quality items. There has been much written about item quality (e.g., most introductory measurement texts include extended sections on item quality) so we do not want to go into detail about this issue here. However, potential users of commercial systems should be encouraged to conduct regular and structured (even expert led) reviews of the quality of items that will be used on their tests. Similarly, those who intend to develop a customized or other local system need to attend carefully to the quality of the items and tasks used to ensure that the assessments can validly fulfill the intended purposes. That is the items should undergo bias and sensitivity review as well as content review. Item should be field-tested whenever possible. These quality checks might require the use of expert consultation to help with local systems. It sounds overly obvious to say that the quality of the interim assessment system is dependent upon the quality of the items included in such systems, but this point often gets overlooked.

Administering the Test

The final consideration in evaluating or designing an interim assessment system is the manner in which the test will be administered. As mentioned early, a key component of many of the commercially-available interim assessments is that they produce results quickly. Most, if not all, commercially available interim systems are computer based, and often web based. The use of technology has several advantages:

- ✓ Allows for flexibility in where the test is taken (anywhere with a computer)
- ✓ Allows for flexibility in when the test is taken
- ✓ Easily creates new forms for different purposes
- ✓ Allows teachers or curriculum leaders to customize forms easily
- ✓ Allows for the possibility of an adaptive test
- ✓ Gives results in real time

- ✓ Aggregates results quickly across a classroom, school, district, state
- ✓ Easily disaggregates results by student subgroup
- ✓ Calculates normative information on the spot

An obvious caveat to the use of the technology is the availability of that technology. If a teacher is to use effectively a web-based system, all students in the classroom need to have access to a computer connected to the internet. If only one such computer exists in the classroom, other provisions may need to be made. Some technology-based systems also include a pencil-and-paper option.

Other potential drawbacks or limitations of being technology-based include:

- ✓ Requires teachers to know how to use the technology effectively
- ✓ Too much customization may reduce the reliability of the assessment
- ✓ May be more difficult to monitor who is actually taking the assessment and under what circumstances
- ✓ Does not allow for interaction between the student and instructor, losing a source of valuable information
- ✓ Current use tends to be limited to selected response items although automated and artificial intelligence approaches to scoring open-ended items are progressing rapidly, particularly in fields such as medicine

Depending on the purpose of the assessment, one should also consider the diagnostic benefits of a teacher administered and scored assessment. Formats such as checklists, performance tasks, and orally administered assessments can provide insights into student thinking that is not readily apparent from a technology-based assessment. Quick turnaround of results is not always essential depending on the purpose. For example, if one purpose is to explore students' thinking and to give teachers better insight into the effectiveness of their instructional strategies, an essay test could provide rich information, even though it may take days, not minutes, to score.

Matching the Administrative Features to the Purpose of the Assessment

The administrative requirements must be considered in conjunction with the test purposes. Depending on the goals of the assessment system, certain features will be necessary, others helpful although not crucial, and other features become irrelevant. Exhibit 2 provides a crosswalk of interim assessment purposes with several administration features.

Exhibit 2: Crosswalk of Assessment Purpose by Administration Requirements

Purpose	Speed of results	Availability of normative information	Flexibility of test administration	Customization of test form	Adaptive
Predict student achievement	Days	Not necessary	Could be given on a flexible schedule or a pre-set times during the year	Not necessary	Nice feature, but not essential
Provide information on strengths and weakness of a particular group of students	Within a few days	Necessary	Needs to be able to be given at the teacher's discretion	Helpful	Nice feature, but not essential
Reinforce curricular pacing	Days-	Helpful	Not necessary	Not necessary	Not necessary
Evaluate the effectiveness of instructional program	Weeks	Necessary	Not necessary	Not necessary	Nice feature, but not essential
Evaluate a student's understanding of a topic	Within a few days	Not necessary	Essential	Helpful	Helpful
Determine a student's preparedness for the next topic	Within a few days	Not necessary	Essential	Helpful	Helpful
Enrich the curriculum	Days or weeks	Not necessary	Essential	Essential	Not necessary
Provide professional development to the teacher	Weeks	Helpful	Not necessary	Not necessary	Not necessary

Summary of Administrative Considerations

In creating this table, we found that we could argue for a different response in some of these cells. For instance, if the purpose of the assessment is to provide exposure to richer curriculum, it could be important to return results immediately, or, if the assessment items themselves provide the instruction, then turnaround time on the results is less important. Likewise, if the purpose of the assessment is to improve a teacher's understanding of the interaction between student learning and performance on the test, it may be important to give the teacher flexibility in determining when to give the assessment and which items to include, or a more standardized assessment could serve the same purpose—particularly if the professional development was intended across a larger scale. However, our purpose in constructing this table was to summarize what was generally the case and to provide an illustration of the thinking an educational leader should do when evaluating or design an interim assessment system.

Evaluative Criteria

To help guide the evaluation of commercially available interim tests, we have provided the following criteria for states and/or district to consider prior to purchasing an assessment system. Following our argument that the interim assessment design must be linked to the purposes and intended uses, we present evaluation criteria for the three major purposes articulated earlier: instructional, evaluative, and predictive. To avoid redundancy, we present several general criteria that cut across all three purposes.

General

1. A test can be no better than the quality of the items it contains. Therefore, the quality of the items needs to be evaluated against professional standards and expert opinion. The types of items/tasks may vary depending on the specific purposes and intended uses, but all should be of high quality as shown through traditional reviews for content and bias and sensitivity as well as pilot testing and data reviews.
2. Alignment evidence should be provided to document the relationship of the items to both the knowledge and skills (including depth of knowledge) called for in the target content standards.
3. The test publisher must include clear guidelines regarding the appropriate uses of the assessment results as well as indicating either potentially inappropriate uses of the results or uses for which there is no validity evidence.
4. Tasks should be applicable to a wide-range of student populations, including English language learners and students with disabilities.
5. There should be evidence that the professional development associated with the assessment system facilitates educators' appropriate interpretation and use of the assessment results for the specified purposes. Clearly, more intensive and sustained professional development is required for assessments serving instructional purposes compared with other purposes.
6. For interim assessment systems that require a break from instruction in order to test, educational leaders should consider the time required for assessment, which should be as short as possible to provide the desired information. For certain performance tasks that are less distinguishable from instruction than more formal tests, the issue of "testing time" is less of an issue, but still must be considered.

Instructional

1. To the extent possible, interim assessments for instructional purposes should be as seamless with instruction as possible and represent an opportunity for student learning during the assessment experience.
2. There should be evidence that the results of the assessment and the associated score reports have been designed to facilitate meaningful and useful instructional interpretations.

3. Clear guidelines should be provided explaining how the results of the assessment, including the results of particular tasks and items, should be used to help inform instructional decisions.
4. Each particular assessment in the system by must be closely linked to the curricular goals taught prior to the assessment administration, preferably quite proximal to the assessment event. The assessment should include only content and skills for which the students have had a legitimate opportunity to learn unless the purpose of the assessment is to determine readiness for some learning in the near future.
5. To best serve instructional purposes, each interim assessment should assess only a limited number of important curricular goals to make it more likely that remediation can be timely and targeted appropriately.
6. In general, interim assessments to serve instructional purposes should be comprised more from high quality open-ended tasks than selected-response items because of the greater probability for correctly diagnosing students' understandings and misconceptions. Multiple-choice items should be developed with an understanding of learning progressions, such that useful information can be gleaned from specific incorrect answers.
7. Instructional interim assessments should measure instructional and curricular goals not easily assessed on the states large scale assessment such as extended tasks or synthesis works.
8. Ideally, the system should provide evidence, based on scientifically rigorous studies, demonstrating that the assessment system has contributed to improved student learning in settings similar to those in which it will be used.

Evaluative

1. The collection of tasks administered through the year should represent a technically sound range of difficulty and appropriate breadth dependent on the focus of the evaluation. Again, this should be examined during the alignment study.
2. The assessments should be comprised of items and tasks with a mix of formats to allow for users to gain a deep understanding of the effectiveness of educational programs.
3. The assessment must be targeted to the content standards that are the focus of the educational program(s) being evaluated or studied and/or to the expected domain of transfer.
4. The reports must be designed to facilitate the intended evaluation. Considering the potential high stakes associated with such assessments (e.g., determining whether or not to spend scarce resources), the reports must accurately portray the error associated with the scores and subscores.
5. The assessment should be related to other measures of the intended constructs and less related to tests of domains other than the intended constructs.

Predictive

1. The assessment should be highly correlated with the criterion measure (e.g., the end-of-year state assessment). The technical documentation should include evidence of the predictive link between the interim assessment and the criterion measure. However, in order to justify the additional testing and cost, the predictive assessment should be significantly more related to the criterion measure than other measures (e.g., teachers' grades) that could be used without adding another assessment to the system.
2. The predictive assessment should be comprised of items with a similar mix of item types as the criterion measure.
3. The predictive assessment should be designed from the same or similar blueprint as the criterion measure, but each test should include only content on which the students have been instructed up to that point.
4. The reports should be designed to facilitate the intended predictions including an honest and accurate characterization of the error associated with the prediction both at the total score and subscore levels.
5. The assessment should contain enough diagnostic information so that remediation can be targeted for students predicted to score below the cut on the criterion measure. If the assessment is unable to provide such information, additional guidance should be included in the system to help with remediation.

We are not suggesting that interim assessment systems must meet all the criteria listed above before being purchased for a district or state, but we suggest that educational leaders consider the criteria when evaluating which, if any, system to purchase or when evaluating a proposal to create a customized system. Additionally, any vendor should be required to provide evidence of the validity of the system for the intended purposes. Once the system has been implemented, districts and/or states should evaluate the system to ensure that it is meeting intended purposes and uses. While any evaluation will have to be tailored to the specific purposes and uses, we offer the following general suggestions for exploring the validity of an interim assessment system:

- ✓ If the test is used for instructional purposes, follow up with teachers to determine how the data were used, if they provided useful information, and whether there was evidence of improved student learning for current students.
- ✓ If the test is used for evaluative purposes, gather data from other sources to triangulate results of interim assessment and follow up to monitor if evaluation decisions are supported.
- ✓ If the test is used for predictive purposes, do a follow up study to determine that the predictive link is reasonably accurate, more than things such as grades and teacher judgments, and that the use of the test contributes to improving criterion (e.g., end of year) scores.

In addition to these suggestions, interim assessment systems should be evaluated for the effects on important aspects of the teaching and learning process, such as:

- ✓ Student learning, especially in terms of generalizability and transfer
- ✓ Student motivation as a result of engaging with these tasks
- ✓ Curricular quality as a result of incorporating tasks
- ✓ Increases in teacher knowledge of content, pedagogy, and student learning
- ✓ Manageability, including the quality of implementation

Current Commercially Available Interim Assessment Systems

As mentioned earlier, many test publishing companies offer interim assessment products, often labeled “formative” or “benchmark” assessment products. Before writing this paper, we searched the internet for companies that offered these types of assessments by entering the terms “formative assessment” “predictive assessment” and “benchmark assessment” into search engines. After reviewing over a dozen websites of various vendors marketing these products⁷, we found many commonalities across the systems.

These assessments are marketed to serve a plethora of purposes, including serving as a diagnostic tool, providing information that can be used to guide instruction, determining student placement, measuring growth or progress over time, and predicting success on a future assessment. Typically these systems consist of item banks, administration tools, and customized reports. These systems often are computer-based and even web-based, allowing students to take the test whenever they wish (or their teacher wishes) and wherever a computer with an internet connection is available. Others also have the option of creating pencil-and-paper tests. Teachers can construct the tests, the tests can be fixed by an administrator, or the tests can be adaptive.

The items are “linked” to content standards⁸, and results typically are reported in terms of number correct. The “diagnostic” portion tends to be a summary of results by content standard, allowing the teacher to see which standards students perform well on and which they do not. Often these systems provide a variety of options for reports, with different levels of aggregation. A student-level report indicates which items students answered correctly or incorrectly, while a classroom report might indicate the percentage of students answering each item correctly or the average percent correct for each content standard.

Some of the products have been linked to state end-of-year assessments, allowing them to serve a predictive function. That is, the student score on the interim assessment is used to predict performance on the large-scale summative assessment. Some of these systems have quantitative data showing the statistical link between the interim and summative assessments

⁷ And sometimes even the same vendor selling multiple versions of their product using all of these labels!

⁸ Unfortunately, the strength of the alignment between such commercial tests and the state content standards is rarely evaluated by independent analysts, so the “link” between the two is often based on the publishers’ claims.

and the correlations between the scores as evidence of the interim assessment's predictive ability. They usually include statistical data on their reliability as well.

These products are marketed as being very flexible, giving instant feedback, and providing diagnostic information on which areas need further instruction. However, these systems generally fail in providing rich diagnostic feedback regarding student thinking. That is, few provide any information on why a student answered an item incorrectly or how best to provide corrective feedback. For instance, many of these computer-based assessments rely primarily on multiple-choice items. Unless each wrong answer provides insight into the nature of the student's incorrect thinking, the only information received from this type of item is essentially a correct/incorrect dichotomous response. Likewise, open-ended items need to result in more than a score, preferably in a summary report of the types of errors a student is making or of the areas of strength and weakness in a given performance (e.g., his/her writing).

In addition, policymakers should consider the validity of these assessments as part of their overall state comprehensive assessment system. The alignment of these assessments to the state content standards should be evidenced through an external alignment study. Furthermore, these assessments should be considered in the wider context of a comprehensive, balanced assessment system. Policymakers should be aware of the potential for several assessments to provide conflicting information. This is particularly problematic when a teacher is using a short-cycle formative assessment in her classroom and receiving different information from a district-wide benchmark assessment. If the short-cycle and medium-cycle assessments are providing conflicting information on the content students have mastered, they are not likely to be useful to that teacher.

In spite of these caveats and concerns, the best current commercially-available systems can:

- ✓ Provide an item bank reportedly linked to state content standards
- ✓ Assess students on a flexible time schedule wherever a computer and perhaps internet connections are available
- ✓ Provide immediate or very rapid results
- ✓ Highlight content standards in which more items were answered incorrectly
- ✓ Link scores on these assessments to the scores on end-of-year assessments to predict results on end-of-year assessments

Many of the better commercially-available interim assessment products can address questions such as:

- ✓ Is this student on track to score Proficient on the end-of-year NCLB tests?
- ✓ Is the student improving over time
- ✓ Which students are at risk of scoring below Proficient on the end-of-year NCLB tests?
- ✓ Which content standards are the students' performing relatively best (or worst) on⁹ (for a student, classroom, school, district, state)?
- ✓ How does this student's performance compare to the performance of other students in the class?

⁹ This assumes that there are enough items for given strands or standards to determine if differences are reliably different.

Although most systems meet some of the requirements for an effective interim assessment, few, if fully meet all of the criteria. Again, the focus remains on the purpose. If the purpose of these assessments is to enrich the curriculum, challenge the students to self-diagnose their own learning, provide insights into any misconceptions the students have, or provide additional professional development for the teachers, many of these types of assessment systems are woefully inadequate.

Thus, we find that most commercially-available interim assessment systems currently do not:

- ✓ Provide rich detail about the curriculum assessed
- ✓ Provide a qualitative understanding of a student's misconception(s)
- ✓ Provide detailed information on the student's depth of knowledge on a particular topic
- ✓ Further a student's understand through the type of assessment task
- ✓ Give teachers the information on how to implement an instructional remedy

Furthermore, these systems cannot answer the following questions:

- ✓ Why did a student answer an item incorrectly?
- ✓ What are possible strategies for improving performance in this content area?
- ✓ What did the student learn from this assessment?
- ✓ What type of thinking process is this student using to complete this task?

Matching the Purpose with the Assessment

The main driver of this paper was to provide advice on how to evaluate the suitability of commercially-available or locally-created products for states and districts considering implementing some sort of interim assessment system. We have continued to emphasize the need to articulate the purpose(s) of such a system. To make this idea more concrete, we have created some hypothetical case studies.

We have assumed different assessment goals for each scenario and mapped possible assessment types using examples of what currently exists along with what we would like to see developed. Each case study explores the desired reporting features, item types, and administrative aspects best suited to meet the specific purpose of the fictional district.

Case Study 1: Exploring a Student's Level of Understanding and Areas for Further Instruction

In this first case study, we have a district that is interested in developing an interim assessment system that can be used in the classroom to help a teacher gauge her students' understanding of what has just been taught and identify areas that need further instruction, either for remediation or enrichment. The goal is not to provide this information to the district policymakers, but results must be aggregated across classrooms to provide periodic reports to the school administrator. Ultimately, the district is not interested in predicting future achievement as its philosophy is that if students are taught the curriculum sufficiently well, they will perform successfully on the end-year-test. Their goal is to develop a system that is fully aligned with the state content standards, will evaluate students' understanding, help the school administrators monitor the students' learning, and provide professional development to teachers on how to administer the system, interpret the results, and apply instructional strategies to remedy any problems found in student learning.

From the reporting perspective, the most important level of reporting is the student level. Teachers need concrete information on what each student knows and does not know, and specifically where the lack of understanding or misunderstanding is occurring. Moreover, the reporting system needs to include information on further probes to explore the student's thinking and suggestions for teaching strategies to effectively counter any existing misconceptions or hindrances to the student's learning. Corrective feedback needs to be provided to students as well as suggestions for further exercises that help the student gain a rich understanding of the concepts covered.

In addition, teachers should have the tools to explore overarching patterns across the classroom in terms of which concepts were clear and which need further instruction. So, classroom-level results should be available. That is, there should be a mechanism in place of quickly aggregating individual student reports to a classroom report. Finally, there needs to be a mechanism for providing this classroom report to a school administrator, along with a summary of any further probes used and instruction given.

For the assessment design, the assessment should be integrated directly into an instructional unit with items linked to specific concepts within a unit. Care should be taken to ensure these concepts (and ultimately the items) are linked directly to the state content standards that will be assessed at the end of the year. Preferred item types are open-ended items that include performance tasks or essays. These items should require critical thinking and student self-reflection. For quick results that allow for easy aggregation, we could consider a system of multiple-choice items, checklists, or short constructed-response items that are followed by probes of why the student thinks their answer is the right answer. These data could be supplemented (or replaced) with a performance task such as asking the students to design and conduct a scientific experiment and chart and analyze the results (science) or developing an argument and writing an essay on why a historical event was inevitable or could have been prevented (history or language arts). Each component of this type of task could be analyzed to see where any misunderstanding or incorrect procedure occurs. These results could be used to give corrective feedback to an individual student and aggregated in a report that shows the percentage of students who can correctly articulate a hypothesis, develop a scientific design with all the appropriate components, record data, graph data, interpret data, and draw conclusions (science) or who demonstrate a rich understanding of a historical event, can develop a persuasive argument, and develop an essay appropriately combining opinion with fact (history or language arts).

Administration of this type of assessment requires much flexibility. The timing may be linked to a date on the calendar or a point in the curriculum. The administration may be computer-based or consist of paper-and-pencil tasks. Even if a teacher intends to read the students' essays or hypotheses, the student can type them into a computer. Results should be available relatively quickly to provide corrective feedback during the course of instruction although certain components, such as essays, may require more time to score. Some further probes may take additional time, but, again, this assessment is intended to work within instruction, not as a separate component.

This type of assessment requires a strong professional development component to be effective. Teachers will need to be trained to evaluate the information and use these types of tasks meaningfully, including modeling how to learn from student work. Professional development supports should also include next steps that reflect the information provided by the assessment

results. Any type of professional development on using assessment results to improve instruction should be continuous, implemented throughout the school year, and not a one-time course.

Case Study 2: Evaluating the Effectiveness of a Curriculum or Instructional Program to Identify Areas for Professional Development

Consider a district with multiple schools, with varying demographics and instructional approaches. The district leaders and policymakers are interested in having information to monitor that schools are providing fair opportunities to learn the required content and skills and to provide information to help evaluate the effectiveness of various district initiatives, including curricular reform activities. The district leaders are interested in receiving more in-depth information about student mastery of specific strands than can be obtained from the end-of-year assessment.

Or, to focus on a more specific simplistic (although less realistic) example with similar requirements, think of comparing two groups of schools, each using a different approach to teach mathematics. One is using a basic skills approach, while another uses an approach that requires integrating mathematical concepts into the everyday world. The first approach has a long list of knowledge and skills to be taught while the second approach has fewer knowledge and skills but requires greater depth of understanding for each. The two schools are similar in terms of student demographics and have performed similarly on previous assessments. The district superintendent wants information throughout the school year on which program appears to be most effective so that information can be provided to the other schools about this instructional approach.

The type of report that will best serve these purposes is a school-level report. The purpose is not to intervene at the student level but to evaluate the approach for the students as a whole. Therefore, the results will need to be aggregated across the school, and then disaggregated by student demographic group to look for any discrepancies in the effectiveness across students. In addition, student subscores would be important, as the policymakers would want to examine the results by the different content strands or instructional units.

The requirements of the assessment design could be similar to those for the end-of-year tests if a policymaker or educational leader wishes to analyze performance towards the goal at several points throughout the year. In this case, the items would need to map directly to the content standards and be similar in type to the items on the end-of-year test to provide information during the year on how prepared students are for the summative assessment. However, additional probes should be added to provide further understanding regarding a student's thinking and solution-strategy for each item. Another approach would be to assess similar domains using different formats to gather more in-depth information to complement the summative data resulting in a more robust evaluation. In this case, the policymakers might consider an open-ended supplement given periodically to provide greater insight into the students' level of understanding.

Finally, the administration requirements should be fairly standardized across the schools. Because these assessments will be used to inform instruction in the next school year, the results do not need to be turned around quickly. Either computer-based or paper-and-pencil tests would be appropriate for this situation. Students can be matrix sampled, too, to provide a

similar level of information with less of a burden on instruction. Measuring achievement periodically of a sample of students from each school will allow policymakers to evaluate the students' learning and growth over the year in the two programs and make appropriate comparisons.

In this case, the designers of an evaluative assessment system must take care to avoid privileging one type of assessment format over another so as not to unintentionally favor one curriculum approach more than the other. By linking carefully to the end-of-year criterion measure, the designers can document that their evaluative assessment is as fair as possible. In the case of this example, the system would likely include only two or three tests spaced widely through the year. This approach is to avoid favoring one curricular approach over another as a result of differential pacing relative to the assessment targets.

Case Study 3: Predicting Success on the End-of-Year Tests and Identifying Gaps in Knowledge

For the final case study, consider a district that wishes to implement an early-warning system to identify which students, classrooms, and schools are on track to perform well on the end-of-year assessment and which might need intervention to meet the annual measurable objective. Furthermore, for those students who are not on track, the district wants to be able to identify areas of weakness at the student level and aggregated to the classroom and school level. Ideally, they would like a system that provides additional tools for improving performance on those areas identified as weak. Finally, they would like to administer this test 3–4 times over the year and track student progress toward the goal.

First, let's examine the necessary components of the reporting system. The first criterion is that the assessment reports "on-track to succeed" as well as any areas of weakness. The results need to be aggregated across classrooms, schools, and the district. Scores should also be disaggregated by the same reporting categories used in the end-of-year reports, such as student racial/ethnic group, gender, disability status, economic status, and LEP. Each subsequent report should illustrate progress, providing feedback on where the student is, how they have developed over time, and how the progress relates to where they should be by the end of the year. That is, the report should show each student's current and project trajectory.

A key component to this reporting system is that areas of weakness for a student, classroom, and school must be clearly identified. Ideally, this information would be deeper than simply identifying content strands or indicators where students were less likely to succeed, but would also focus on specific concepts and provide follow up activities for each potential area of weakness. It should include instructional strategies based on research for improving student performance.

Next, let's look at the requirements of the assessment design. The items need to map directly to the state content standards and be similar in type to the items on the end-of-year test to provide a solid statistical link for predictive purposes. However, each test should only assess what's been taught to date. That is, this system should not be designed to give a series of parallel mini-summative assessments. The first assessment should cover the material that was scheduled to be taught in the first quarter, for example. The second assessment may provide some overlap with the material taught toward the end of the first quarter and cover the material taught in the second quarter. It is also important that the items link not only to the

state content standards but to teaching units and text books specific to that district. Using items that link directly to instructional materials will help provide the connection between any weaknesses found and instructional interventions.

And although it is important to use similar item types as used in the end-of-year assessment for statistical reasons, consideration should be given to adding a few open-ended probes to help diagnose any weakness found in student performance and to allow these less constrained assessments to measure student performance in ways unable to be assessed on the single end-of-year tests.

Finally, let's consider administration requirements. Because these assessments are not designed to integrate seamlessly with instruction but rather as a periodic check, it is not as important that the results be turned around within a day. However, the results should be available within a week or so to allow time for intervention. Particularly assuming that one unit builds on the next, it will be important to inform teachers quickly if students are not on track to succeed.

In terms of the actual administration, either computer-based testing or a pencil-and-paper test would serve this district's purpose. Also, standardization in the items administered would be necessary to aggregate results across the district. Therefore, flexibility is not a strong requirement for this system. Furthermore, since the goal is to ensure all students are on track to meet a certain criterion, normative data are not as important to provide as criterion references.

A critical assumption justifying this predictive use is that the results provide useful information beyond what teachers already have. If the results of predicting which students had mastered a unit sufficiently to be successful on the end-of-year assessment were quite different from the teacher's predictions or grades for the student, then there would be concerns about the degree of mismatch between the instruction and the end-of-year assessment. Likewise, if the assessment provided no additional information beyond what the teacher already knew from her daily interactions with the class, then the assessment would not be worthwhile.

Summary of the Case Studies

We recognize that most actual instantiations of interim assessment systems do not map perfectly onto one of these cases. In almost all cases, educational leaders are trying to squeeze as many purposes as possible out of a single system. Unfortunately, one of the truisms in educational measurement is that when an assessment system is designed to fulfill too many purposes—especially disparate purposes—it rarely fulfills any purpose well. This does not mean that certain interim assessment systems cannot fulfill more than one purpose, depending on the level for which the primary purpose is intended to address. If the system is intended to provide rich information about individual students' strengths and weaknesses tied to a particular set of curricular goals, then these results can likely be aggregated to the subgroup, school, and/or district level to provide evaluative and predictive information. On the other hand, if the primary goal is to gather predictive or early warning information, it is unlikely that the assessment will contain rich enough information for full instructional or even evaluative purposes. Therefore, if users want to fulfill multiple purposes, they must design a system to fulfill the finest grain purposes first and then consider approaches to aggregate the results to more general levels in the educational system.

Discussion

Throughout this paper, we have attempted to maintain a neutral tone while providing an overview of interim assessment systems. Now, however, we wish to use this discussion section to express some of our concerns with the current use of interim assessments and our hopes for the future direction of this work.

We first approached this paper from the perspective of promoting formative assessment. However, as we examined what is now in the field under the appropriated term “formative assessment” we realized that there needed to be a discussion regarding the current types of assessments being sold for formative purposes. When asked why we chose to focus on interim assessments rather than the purer formative assessment, our answer was simple: because states and districts are spending considerable resources to implement such systems. We recognize that to develop a strong formative assessment system as advocated by Black, William, Shepard, and others is difficult to do at a state level. Components such as weaving the assessment seamlessly into the curriculum and providing useful feedback that leads to appropriate modifications in instruction is difficult when the agent (state DOE personnel) is several steps removed from the classroom. While states can support professional development programs that teach our educators how to develop and use such tools, it would be helpful if states and districts could purchase a pre-existing system that supports formative and professional learning needs. In addition, states may have other requirements for an assessment program, such as to develop an early-warning system to identify students who are not on track to succeed and give them additional supports. Or, the states may wish to use these interim assessments as evaluation tools for different schools, instructional programs, or pedagogy. That is why we chose to define interim assessments as tools to evaluate students’ knowledge and skills relative to a specific set of academic goals that are designed to inform decisions at a level above the classroom and to focus our discussion on the purposes and uses of such assessments.

That said, we are concerned that many of the commercially-available systems have moved far a field from what the research currently supports and those selling such system promise more than they can deliver. For example, these systems often lay claims to the research documenting the powerful effect of formative assessment on student learning when it is clear that the Black and William meta-analysis evaluated studies with formative assessments of very different character than essentially all current commercially-available interim assessment programs.

We believe that an interim assessment system that simply administers a series of mini-summative assessments is ineffective. We have seen several systems where shorter versions of the end-of-year assessment are given periodically. The items on these assessments are placed on the same scale as the items on the end-of-year assessment, so the results can be used to show progress towards the goal. One system we saw gave essentially the same form up to four times a year so the students actually saw the same or very similar items multiple times. We believe these systems are not the best use of money or instructional time. Testing students on material they have not yet been exposed to is useful only when one is unaware of what they have and have not been taught. Why test students in October on the material scheduled to be taught to them in February? What have you learned when you see they do better on the same items in March?

A good interim assessment can be an integral part of a state's comprehensive assessment system, used in conjunction with classroom formative assessments and summative end-of-year assessments. As such, we believe that there are valid purposes for giving interim assessments beyond informing instruction at that point. However, the policymakers and educators using the assessment need to understand the purpose of the assessment and what it can and cannot do. If policymakers want an assessment to help educators improve instruction, they should look for one that ties directly to the classroom instruction and provides in-depth examination of not just which items students miss but why they miss them. And, most importantly, they need to provide the next step—an intervention for correcting a misconception or teaching a missed lesson. If policymakers want an assessment to tell them how students are likely to perform on an end-of-year assessment, they need to examine the reliability of the predictions and the information describing what to do next. Is the purpose of the assessment simply to identify students not on track in order to place them in remedial instruction or provide additional supports, such as tutoring? Or is it the teachers' responsibility to interpret the results of the tests, determine areas of weakness, and correct them within the classroom? The requirements of the assessment would be different depending on the desired next steps.

At a minimum, we argue that any expenditure of resources (teacher time, money, etc.) for an interim assessment system must provide experiences and information that is not available on the state large scale assessment or in the classroom through daily instructional activities. This additional information would include concepts such as allowing for the assessment of deeper learning that is challenging to address on the large scale assessment and more information about particular content strands than can be accomplished on typically short teacher-created tests. Finally, any of these assessment types need to provide evidence of their validity. Are they demonstrating their intended positive consequences and are there any unintended negative consequences of their use? For instance, do additional assessments solidify a student's understanding of a concept or inure them to tests in general? These interim assessments are an integral part of any comprehensive assessment system and should be considered as a piece of a whole and evaluated as such.

There are organizations trying to develop the types of thoughtful probes we discussed, but there are others that are simply trying to sell item banks and reporting systems. Our goal here was not to condemn all currently available products, but rather to provide a framework for the consumer to use in evaluating them. An additional purpose of this paper is to promote interest in further research in this area, and to that end, we conclude by describing our vision for this research.

Future Areas of Research Needed

This field is rich for further research. New studies funded by the U.S. Department of Education's Institute for Education Sciences (IES) are exploring areas that may serve to inform the field of formative uses of assessment. Many of these studies focus on interim assessments, sometimes as part of a tutoring session or computer-based learning. In general, they examine how testing a particular unit of instruction was related to retention of information after an extended period of time. One common finding across studies was that student performance on the "repeated testing" was not nearly as important as the corrective feedback they received as a result. That is, a student who guessed incorrectly on an item on a unit test, but who received good corrective feedback was just as likely to answer a similar item correctly on a future test as a student who had answered it correctly the first time. Another common finding we found

interesting was that the repeated testing, in and of itself, contributed to retention. And this was particularly true when the short tests required students to generate their own responses on short-answer items (Viadero, 2006). We look forward to seeing the results of these studies published.

We feel it is important to continue to examine how the use of interim assessments can help further student learning. In particular we see the need for research in the following areas:

1. Create a validity argument for how interim assessments lead to success on summative assessment and gather evidence to evaluate this argument. Choose several types of interim assessments and validate their uses.
 - a. Is a predictive assessment truly predicting student performance on an end-of-year assessment more so than other readily available data? Of course, the results of this question could be confounded by the use of appropriate interventions, but those interventions may provide evidence of the validity of the consequences.
 - b. Is the use of instructional assessments improving instruction? Are there any unintended consequences?
 - c. Are evaluative interim assessments effectively identifying differences in various pedagogies or instructional approaches?
2. Examine differential effects of interim assessments on students' intrinsic motivation to learn. Consider the theory that frequent assessments can diminish intrinsic motivation by shifting the effort and purpose from learning "to know" to learning so as "to display one's knowledge" (Lave & Wenger, 1991). How can we use the interim assessments constructively to further students' desire to learn rather than to further their desire for a high score?
3. Examine the types of feedback that are most effective for improving student performance. Kluger and DiNisi (1996) found that normative types of feedback or feedback that focuses on the person rather than on the task can actually have a negative effect on student performance. Their research showed that the most effective types of feedback were ones in which students were told not only what they needed to learn but how to get there. How does this research apply to the interpretation of results from interim assessment?
4. It has been argued that evidence collected for summative purposes can rarely be disaggregated to support learning, but evidence collected for formative purposes can be aggregated to support summative inferences (William, 2006). However, we need to learn more about how to aggregate results of formative assessments. What are the requirements for building a system that provides teachers the information they need but can still be scaled to compare results across students, teachers, and/or schools?
5. What types of professional development are linked to effective use of interim assessments? What is the best delivery system for this professional development?

References

- Black, P., Harrison, C., Lee, C., Marshall, B., & William, D. (2002). *Working inside the black box: assessment for learning in the classroom*. London, UK: NFER-Nelson.
- Black, P. & William, D. (1998). Assessment and classroom learning. *Educational Assessment: Principles, Policy and Practice*, 5(1), 7-74. Also summarized in an article entitled, Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-148.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254-284.
- Lave, J. & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.
- National Research Council (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy of Sciences.
- No Child Left Behind Act of 2001, Pub. L. No.107-110, 115 Stat.1425 (2002).
- Shepard, Lorrie. (June 2006). *Can Benchmark Assessments Be Formative?: Distinguishing Formative Assessment from Formative Program Evaluation*. Presented at the CCSSO Large Scale Assessment Conference, San Francisco, CA.
- Scriven, Michael. (1967). The methodology of evaluation in perspectives of curriculum evaluation. In R. W. Tyler, R. M. Gagné, & M. Scriven (Eds.), *Perspectives of curriculum evaluation*, 39-83. Chicago, IL: Rand McNally
- Viadero, Debra. (2006). Cognition studies offer insights on academic tactics: U.S.-funded projects eye ways of helping students remember more material. *Education Week* August 30.
- William, D. (2006). Assessment for learning: why, what and how. *Orbit: OISE/UT's magazine for schools*, 36(3).

Appendix 4

Other State Reports

State Assessment Task Forces								
State	Year	Audience	Members	Report Focus	Resources	Recommendations	Task Force Final Report	Teacher Resources
Colorado	2015	General Assembly	15: representing parents, teachers, administrators (various districts/charters), school board (2), businesses (Chamber of Commerce, education CBOs). Additional online public commentary and community feedback meetings held.	To study implications of statewide assessment system from the ground up, report findings, and present legislature with short term, actionable feedback.	<p>Colorado Assessment Implementation Study: assess concerns with implementation of new tests and provide feedback to inform policy, practice, and future direction specifically concerning time, readiness, quantity, frequency, length, need for timely/relevant results, burden, utility, and types.</p> <p>Study of Assessment Use in CO Districts and Schools: Assessment timelines, cost, preparation, existential costs to classroom education, benefits/impacts, suggested changes.</p> <p>Background info on assessments/practices and response to questions: CDE</p> <p>Stakeholder input: focus groups, online surveys, and hearings.</p>	<p>Data driven policy implications.</p> <p>Robust</p> <p>Aligned with current TN assessment practices and areas of interest.</p> <p>Although recommendations will directly impact classrooms, the report lacks education-specific best practices.</p> <p>Future policy considerations are included in report to address long-term issues.</p>	https://www.cde.state.co.us/cdedepcom/finalreport1202taskforce	<p>Additional teacher resources for state assessments include curriculum and unit examples, webinars regarding development.</p> <p>https://www.cde.state.co.us/standardsandinstruction</p> <p>Assessment information</p> <p>https://www.cde.state.co.us/assessment</p>

State	Year	Audience	Members	Report Focus	Resources	Recommendations	Task Force Final Report	Teacher Resources
Iowa	2014	General Assembly	21: teachers, administrators, tech assistants/prof development provides, higher ed, 1 DOE rep, Iowa Business council, 1 parent	To evaluate assessment needs (implementation related) and recommend a new state assessment for schools.	Various assessments from vendors Scoring Rubrics of Initial Assessment Review, provided by assessment vendors Vendor Q&A	Report focused on implementation of tests and, although does not provide best practices in regard to teaching, does address relevant issues in relation to administration and roll out, as well as considerations for assessment-standards alignment through student progress monitoring.	https://www.educateiowa.gov/sites/files/ed/documents/2014-12-31AssessmentTaskForceReport.pdf	Student Assessment- Tips (general recommendations for improving instruction and assessment) and information: https://www.educateiowa.gov/student-assessment Alignment: https://www.educateiowa.gov/sites/files/ed/documents/Alignment%20of%20the%20Iowa%20Tests.pdf NCLB information: https://www.educateiowa.gov/pk-12/no-child-left-behind/standards-benchmarks-grade-level-indicators One Superintendent blogged in regard to the task force outcomes: http://www.west-branch.k12.ia.us/an-overview-of-the-iowa-assessment-task-forces-recommendations/
Illinois	2015	General Assembly	21: 2 parents, teachers, administrators, 1 board member, assessment expert, SBOE, superintendent, representatives from 2 & 4 year higher ed institutions, 1 senator, and CBOs	Reviews the content and design of standardized assessments; the time and money expended to prepare for standardized assessments; and parent, student, and educator perceptions of the level and intensity of standardized assessments given in local school districts.	Individual surveys were created for superintendents, teachers, parents, and students and distributed through a representative sample of districts via email.	Provided feedback on educator/administrator and community level climate surrounding assessments in both quantitative and qualitative form.	http://www.isbe.net/ARTF/pdf/artf-ga-report-2015.pdf	Assessment resources: http://www.isbe.net/assessment/htmls/teacher-resources.htm Information cheat sheet: http://www.isbe.net/assessment/pdfs/parcc/2015/stds-asmt-resource-1214.pdf