INTRODUCTION:

After the announcement of the Race to the Top Assessment Awards, two consortia submitted applications for funding. The Partnership for Assessment of Readiness for College and Careers (PARCC) and the SMARTER Balanced Assessment Consortium (SBAC) both advanced ideas to revamp the assessment model currently used to measure academic performance in America’s schools. While North Carolina has chosen to join the SBAC, both consortia have included through-course assessments, also called distributed summative assessments, in their plans. This paper will:

• Describe the tenets of this assessment model,
• Discuss its strengths and limitations,
• Give examples of two prominent versions under discussion, and
• Explore the implementation of this assessment model in North Carolina.

THE TENETS OF THROUGH-COURSE ASSESSMENT:

• Academic objectives are divided into three to five units of instruction.
• Students take assessments on intra-year curriculum units.
• Unit results are aggregated to produce a summative score.

North Carolina currently requires a cumulative summative test at the end of each school year for children grades three through eight and at the end of certain courses in grades nine through twelve (North Carolina Department of Public Instruction, 2008). These scores allow the state to gauge Adequate Yearly Progress as dictated by No Child Left Behind (North Carolina Department of Public Instruction, Determining AYP, 2009), as well as to measure academic growth and proficiency as called for by the ABCs accountability model (North Carolina Department of Public Instruction, Evolution of the ABCs, 2009). The End-of-Grade and End-of-Course exams attempt to capture a child’s academic performance across a grade or course in a single test given at its conclusion. Through-course assessment (TCA), on the other hand, tests the student individually on three to five units of academic standards and aggregates the scores into a summative result (Nellhaus, 2010). For example, the state would divide objectives from the newly revised standards for Biology into three units. At the end of each unit, students would be tested on skills and knowledge from that unit. At the end of the course, the three unit exam scores would be aggregated and used in place of an End-of-Course exam score. The state may then use results from these tests for accountability purposes, just as it does with the cumulative summative system.
Concurrent developments in TCA and computer-adaptive testing may go hand-in-hand when envisioning a TCA system. Computer-adaptive testing provides a more accurate picture of student performance by tailoring question difficulty on the basis of success on the previous questions (Rabinowitz, 2010). A recent report to the North Carolina State Board of Education discusses the implications of moving the state to a computer-based (online) testing system (North Carolina Department of Public Instruction, Transitioning to Online Assessment, 2010). It should be noted, however, that the PARCC plans to use TCA independent of a computer-adaptive testing program.

Supporters and detractors of TCA have argued a number of advantages and disadvantages when compared to the traditional cumulative summative method. The following discussion will explore those topics without delving into the issue of implementing a computer-adaptive testing system, as the strengths and challenges therein are not unique to TCA and are covered in the report mentioned above.

**THE STRENGTHS OF THROUGH-COURSE ASSESSMENT:**

The major strengths of TCA lie in its ability to provide both comprehensive information on student mastery and critical formative feedback to inform instruction throughout the school year.

1 The End-of-Course exams are untimed, although the Department of Public Instruction does estimate that it will take 95% of students less than one hundred and fifty minutes to complete an exam. Students are allowed up to four hours to complete an exam, with additional time provided for those with testing accommodations (North Carolina Department of Public Instruction, End-of-Course Items, 2010).

**COMPREHENSIVE ASSESSMENT:**

- Teachers can gauge student performance on specific skills and knowledge.
- Performance opportunities allow for better testing of student mastery.

One of the advantages of holding multiple unit-based testing sessions lies in the opportunity to test more comprehensively. For example, End-of-Course exams include between sixty-eight and one hundred questions and are one hundred and fifty minutes in duration. In that time, the items must test proficiency across material of an entire course. Therefore, proficiency in each content area is, by necessity, assessed with a small number of items. As a result, critics of such exams argue that the assessments cannot provide a reliable measurement of mastery of all different content or skill areas covered within a course.

If each unit test in TCA is given in the same amount of time as the cumulative summative exam, the assessment can more comprehensively cover the content area. Since the tests cover a smaller amount of material, the time allotment may also allow for the inclusion of innovative (non-multiple-choice) items such as constructed response and performance tasks, which require students to demonstrate higher-order thinking and synthesis of skills and knowledge covered in individual units. Such a test is able to assess core standards more deeply than a simple cumulative summative assessment (Domaleski and Hill, 2010).
Moreover, the inclusion of performance tasks, constructed responses, etc. engages the teacher and allows her to see more fully the relationship between core academic standards and her students’ work in the classroom. In short, testing more frequently, using a variety of items, on smaller content clusters helps the teacher see the linkages between standards, instruction, and assessment (Wilson and Sloane, 2000).

FORMATIVE INFORMATION:
- Intra-year assessment informs instruction modification.

A key weakness of the current assessment system is that teachers receive student scores too late in the year or course to alter their teaching styles or focus on areas that suit their students’ needs. The usage of interim assessments in the TCA model gives teachers the opportunity to see their students’ achievement on standardized tests and modify their instruction to provide targeted remediation to low-performing students and offer more challenging material to high-performing students (Wilson and Sloane, 2000).

Proponents of TCA argue this model of assessment more accurately reflects the natural learning progression, particularly in skill-oriented subjects like reading and math (SMARTER Balanced Assessment Consortium, Theory in Action, 2010). Rather than testing a year or course’s worth of content at one time, TCA uses testing to measure content clusters in the order they are learned. Since each successive cluster builds upon the previous one, TCA offers students and teachers information to identify weaknesses in fundamental areas before moving to more advanced topics.

THE LIMITATIONS OF THROUGH-COURSE ASSESSMENT:

TCA is not without its weaknesses. Testing logistics can be more difficult under a TCA system. Additional testing may increase anxiety among students, and it is possible that TCA will actually decrease student retention of all material in a class. Lastly, states need to become more involved in curriculum design, especially the pacing of courses, in order for TCA to function properly.

TESTING LOGISTICS:
- System requires additional time and resources to test and score students.

Moving from a single standardized test administration to three to five within a year or course has a considerable impact on testing logistics (Resnick and Berger, 2010). The current resources required for End-of-Grade and End-of-Course exams would have to be multiplied to accommodate TCA. TCA demands extra overhead in scheduling as it takes additional time for a school’s testing coordinator to schedule multiple exams for each state-tested class. Additional physical and human resources may also be needed to ensure that there are enough testing materials for students and that there are ample proctors for the exams. All of these resources come at a significant cost to the state and school districts.

Since TCA advocates for the use of testing items beyond multiple-choice questions, it will require more human grading than do current summative tests. While artificial intelligence is becoming more reliable in scoring constructed response items, it has not yet reached an optimal stage of development. Grading performance tasks is still somewhat beyond our current technology and will necessitate much human labor, incurring additional costs to pay for those individuals who grade the performance tasks.

NEGATIVE STUDENT IMPACT:
- More frequent high-stakes exams may generate more test anxiety.
- The lack of cumulative testing may discourage information retention.

While TCA may have definite advantages in terms of student outcomes, there are also challenges involved when asking students to take three to five state-mandated tests during the year, particularly if those tests are high-stakes for the students. Clearly, more time spent on testing leaves less time for instruction. Moreover, TCA may well multiply instances of test-stress for students, especially those students who do not test well. The North Carolina General Assembly recently considered an amendment to the 2010 budget to eliminate most tests not required by No Child Left Behind precisely for these reasons (Senate Bill 897, 4 June 2010). As such, there may be political opposition with regard to an increased number of exams.

In a TCA model that does not include a cumulative summative test at the end of the grade or course, there is the possibility that students will retain material only long enough to pass the distributed exam (Wise, 2010). In other words, if a student tests on the first unit of standards early in the year and is not reassessed on those standards before the end of the course, there is less pressure for him to retain that information. This concern is particularly true for subjects in which content and skills do not build on each other. In essence, without a reinforcing mechanism like a comprehensive End-of-Grade or End-of-Course exam, TCA may be a test of students’ short-term memory more than mastery of the content.

INCREASED STATE INVOLVEMENT IN CURRICULUM:
- Model necessitates state-dictated order and pacing of content instruction.
- Schedule may limit curricular flexibility.

In order for TCA to tightly align with Common Core State Standards and allow for meaningful comparisons between schools and local education agencies, as well as across time, the state needs to be highly involved in setting the content clusters and pacing for instruction, as well as the scheduling of each exam (Nelhaus, 2010). This intervention is unprecedented in North Carolina, which has traditionally allowed school districts to set curriculum content and pacing.
Under TCA, educators must teach content in a pre-established order. Currently, the Standard Course of Study sets out objectives to be covered, but not an order in which they must be taught. However, allowing teachers to test their students on different units at different times compromises test security under TCA, and it may lead to a skewing of scores. For example, an English teacher may decide to teach objectives related to poetry at the beginning of the year while another teacher ends the year with the poetry unit. The scores on the distributed assessment for poetry may be higher for the students of that second teacher simply because they have received general reading instruction during the year. As a result, with TCA, teachers will lose some of their freedom to make decisions on curriculum.

Additionally, the timing requirements of this system may diminish the opportunities teachers have to remediate students on prior units or years of content before the next test. As an example, Algebra II students in a high-performing class may begin the school year ready to learn the course’s content and skills. However, in a low-performing class, the teacher may need to remediate on content from Algebra I before beginning the more advanced material. A situation can arise in which students in the low-performing class fall behind schedule right as the class begins.

**THROUGH-COURSE ASSESSMENT IN PRACTICE:**

It should be noted that many teachers already use a system of unit exams and benchmark assessments to measure student progress throughout the year. TCA shifts this practice from the classroom to the state level. No state in the nation has implemented a fully developed TCA system for accountability purposes. The nearest in terms of testing schedule is Oregon’s Assessment of Knowledge and Skills (OAKS), which offers the equivalent of a cumulative summative test a number of times throughout the year. The highest score received then becomes the student’s grade for that class (Wise, 2010). This system is fundamentally different from TCA in that the entire period’s content is tested each time, sometimes before the material is covered in class. The weaknesses of our current system with one cumulative summative exam are still present in the Oregon model, except that students and teachers can benefit from the formative use of results from earlier testing.

**POPULAR MODELS FOR THROUGH-COURSE ASSESSMENT:**

As the demand for more innovative, performance-based assessment has increased, two models for TCA have elicited the most discussion and attention. Linda Darling-Hammond and Ray Pecheone from Stanford University have designed a “Balanced Assessment” system, while Lauren Resnick from the University of Pittsburgh and Larry Berger from Wireless Generation have collaborated to create the “American Examination System.” Both models utilize technology and more frequent assessments to better examine the depth of student mastery of standards.

**DARLING-HAMMOND AND PECHEONE’S “BALANCED ASSESSMENT” SYSTEM:**

In the “Balanced Assessment” system, teachers would administer performance tasks to students throughout the course. At the end of the course, students would take a cumulative summative exam that includes some performance-based items, such as short-answer questions. Computer-adaptive testing would be used for the summative exam, while the use of rubrics would guide human and artificial intelligence programs in the grading of performance tasks. The student’s final summative grade could be determined by weighting the results of the performance tasks with the summative exam score (Darling-Hammond and Pecheone, 2010).

**Strengths of the “Balanced Assessment” System:**

- States can choose to decrease testing duration or test more deeply.
- The cumulative summative exam tests understanding of entire course.

As with any assessment models that utilize computer-adaptive testing, students may need to answer a smaller number of test items to pinpoint their degree of mastery. In more traditional pencil-and-paper assessments, students take an established number of test items, many of which may be too simple or challenging to provide information on their true achievement level. Since some skill levels require fewer testing items to ascertain than the current system employs, the time required for a testing session decreases when using computer-adaptive testing. Alternatively, tests may utilize the customary amount of time and assess more deeply those content areas within the unit. However, an expansive item bank must include questions at various difficulty levels for all standards. States can share the cost of the development of such an item bank by collaborating with each other. Estimates on the cost of a “Balanced Assessment” system are $10 - $20 per pupil depending on how exams are scored. This amount is less than many state testing programs (Rabinowitz et al, 2010), although North Carolina estimates that the administration of state exams costs approximately $15 per pupil.

The “Balanced Assessment” model also counters one of the main critiques of TCA: that students are only tested on the most recently learned information. Critics argue that each distributed assessment’s focus on the past unit of instruction neglects to assess students on the total amount of skills and knowledge.

---

2 State exams in North Carolina vary in cost; the average of the per-pupil cost for each test is $15.18. Costs for End-of-Grade and End-of-Course exams, as well as alternate assessments (such as NC EXTEND), were used to determine this average.
that they have accumulated (Nelhaus, 2010). Darling-Hammond and Pecheone’s model includes a cumulative summative exam to better reflect what students have learned throughout a course, not just during that last unit of instruction.

**Weaknesses of the “Balanced Assessment” System:**
- The summative exam does not provide formative information.

The summative exam in the “Balanced Assessment” system cannot be used for formative purposes, including changes to teaching style or additional time spent on challenging topics. Information gleaned from student test scores will not be used to change instruction until the following school year, when it is already too late to help the current students (Kennedy et al, 2005). Therefore, advocates offer, the summative assessment is an unnecessary component to the system.

**RESNICK AND BERGER’S “AMERICAN EXAMINATION SYSTEM”:**
In the “American Examination System,” content standards for a class would be divided into three to five units of instruction. At the beginning of each unit, students would take a pre-test. Teachers would then administer an exam at the end of the unit, or whenever a teacher decides that a student is prepared to test. The score on the exam would serve as a post-test to measure student growth. At the end of the year, the scores on the post-test exams would be aggregated to produce a summative exam score. Students would have multiple opportunities to pass the distributed exams. There would be no cumulative summative exam (Resnick and Berger, 2010).

**Strengths of the “American Examination System:”
- Teachers can assess student growth at the unit-level.
- If tests limit performance items, quick grading yields formative feedback.

As with most TCA models, a major benefit is that more frequent testing allows for comprehensive assessment of the standards. It is also possible to measure growth at the unit-level through the use of pre-tests and post-tests. This model more closely follows the natural progression of student learning, that we discussed above (Kennedy et al, 2005). Psychometricians can validate the pre- and post-tests to ensure that the exams assess desired content (for example, the Common Core of State Standards), and to ensure that good teaching increases exam scores. In this validation process, students receive high-quality

---

**FIGURE 1 – The “Balanced Assessment” System**

- Beginning of School Year
  - Unit One of Course Completed.
    - Students complete performance task on Unit One material.
  - Unit Two of Course Completed.
    - Students complete performance task on Unit Two material.
  - Unit Three of Course Completed.
    - Students complete performance task on Unit Three material.

Students take a computer adaptive cumulative summative assessment. The student's final summative assessment grade can be his score on this exam, or a weighted average of his score on the cumulative exam and the performance tasks.

**FIGURE 2 – The “American Examination System”**

- Beginning of School Year
  - Unit One Pre-test
    - Students take distributed accountability exam on Unit One material.
    - Unit Two Pre-test
  - Unit One of Course Completed.
  - Unit Two of Course Completed.
  - Unit Three of Course Completed.

Students take distributed accountability exam on Unit Two material.

There is no cumulative summative final exam. Summative exam score determined by aggregating scores on Unit One, Two, and Three distributed accountability exams.
instruction focused on academic standards. As a result, they should score higher on the post-test than on the pre-test. If they do improve, psychometricians have good evidence that the tests are valid (Rabinowitz, 2010).

Teachers and students quickly receive feedback from distributed exams, which allows teachers to adjust instruction, and students and families to respond appropriately to scores. Having a more open testing calendar allows for additional time to score constructed-response items. The cost of scoring is what leads to some variation in estimates of the cost of the “American Examination System.” However, estimates are $15 - $20 per student, which is less than many current state testing programs (Rabinowitz, 2010), but close to what North Carolina currently spends for multiple-choice assessments.

**Weaknesses of the “American Examination System:**

- Model contains no cumulative component.
- System’s large test bank and multiple administrations will increase costs.

This TCA model does not include a final exam of any kind. Unless some type of summative exam is incorporated into the system, states run the risks that students will not retain information longer than the length of one unit of instruction, or, if they do, that teachers will not know it (Wise, 2010). Additionally, the costs of designing a larger test bank and administering exams more frequently are greater than the costs with a traditional summative exam testing model. Moreover, this system is vulnerable to the same gaming issues associated with all pre-test/post-test models: teachers may discourage students from performing their best on pre-tests in order to keep those scores low and improve the appearance of student growth.

**Potential Modifications to Both Models:**

- Models may include intra-year or summative cumulative components.

The major concern about both the “American Examination System” and the “Balanced Assessment” model is that they do not place enough emphasis on student retention of a course’s worth of material. As a result, some call on Resnick and Berger to add a summative assessment to measure student mastery of all content standards covered during the year. Students’ final summative grade would then be an aggregation of their summative assessment score, as well as their scores on the distributed accountability exams (Figure 3). This practice will raise methodological questions as to how the scores should be combined to form the student’s “true score” for the year.

Another proposal is to make both Resnick and Berger’s distributed accountability exams and Darling-Hammond and Pecheone’s performance tasks cumulative so that students are constantly reassessed on older standards (Figure 4). This modification is also designed to increase student retention of material (Wise, 2010).

**IMPLEMENTATION OF THROUGH-COURSE ASSESSMENT IN NORTH CAROLINA:**

The TCA model fits well with North Carolina’s current curriculum and accountability reform effort, as well as its work with the SBAC. However, there are major obstacles that must be overcome if TCA is to be implemented in North Carolina.

**EXISTING EDUCATION INITIATIVES:**

- The ACRE initiative focuses attention on use of formative information.
- SBAC benchmarking requires creation of content clusters.

Two years ago, the Department of Public Instruction began the Accountability and Curriculum Reform Effort (ACRE) to re-design the state’s Standard Course of Study, testing program, and accountability model. When fully implemented, ACRE will result in newly focused academic standards and enhanced assessment types, including the use of computer-adaptive testing and performance-based tasks. ACRE also includes a formative assessment component that will offer...
computer-adaptive benchmark exams to provide information on student mastery and professional development tools for teachers on how to use the data to guide instruction. As a result, the use of technology and focus on formative information in the TCA model fits well with ACRE’s current efforts (North Carolina Department of Public Instruction, Accountability and Curriculum Reform Effort, 2010).

This spring, North Carolina joined the SBAC, a group of states collaborating on a Race to the Top grant for assessment design. The Consortium has proposed an assessment system that includes benchmark exams on discrete units of standards from the Common Core of Standards (SMARTER Balanced Assessment Consortium, Master Document, 2010). If the Consortium receives funding, North Carolina will need to move toward the design of standardized units to be used for benchmark assessments. If such units are already in place, the shift to the TCA model will be much smoother.

**POLICY OBSTACLES:**

There are three major obstacles to the move to TCA: current testing policy, opposition to increased government intervention, and scheduling difficulty.

**Current Testing Policy:**

- TCA implementation would require a change in assessment policy.

The State Board of Education and the Office of Administrative Hearings would need to change the policy on assessment of students. Currently, the State Board of Education and General Statutes require that students take End-of-Course exams within five (for a block schedule) or ten (for a traditional schedule) days of the end of the class. This policy would require modification to allow for testing throughout the course.

1. The State Board of Education would first change the policy; it would then be sent to the Office of Administrative Hearings for further action.
2. The Rules and Review Commission would then investigate the impact of the proposed change.
3. If the Commission approves the rule, it would then be entered into code, unless ten or more members of the Commission objected to the rule. If so, it would require legislative approval before becoming final (North Carolina Office of Administrative Hearings, 2008).

---

**FIGURE 4 – Part A: Cumulative “American Examination System”**

- **Beginning of School Year**
  - Unit One Pre-test
  - Unit One of Course Completed.
    - Students take distributed accountability exam on Unit One material.
    - Unit Two Pre-test
  - Unit Two of Course Completed.
    - Students take distributed accountability exam on Unit One and Two material.
    - Unit Three Pre-test
  - Unit Three of Course Completed.
    - Students take distributed accountability exam on Unit One, Two, and Three material.

- **End of School Year**

There is no cumulative summative final exam. Summative exam score determined by aggregating scores on Unit One, Two, and Three distributed accountability exams.

**FIGURE 4 – Part B: Cumulative “Balanced Assessment” System**

- **Beginning of School Year**
  - Unit One of Course Completed.
    - Students complete performance task on Unit One material.
  - Unit Two of Course Completed.
    - Students complete performance task on Unit One and Two material.
  - Unit Three of Course Completed.
    - Students complete performance task on Unit One, Two, and Three material.

- **End of School Year**

Students take a computer adaptive cumulative summative assessment. The student’s final summative assessment grade can be his score on this exam, or a weighted average of his score on the cumulative exam and the performance tasks.
Given the current backlog on rules pending at the Office of Administrative Hearings, it would take longer than six months after approval by the State Board of Education to complete the process.

Political Opposition to Increasing State Involvement:
- School personnel may be opposed to increased state involvement.

As discussed above, TCA requires that the state, or a group of states such as the SBAC, agrees to the division of academic standards into testable units. While North Carolina has long provided standards for core academic subjects, the state has not become involved in the order in which the standards are taught or the pacing of instruction. State extension into these areas is likely to result in political opposition from some teachers, principals, district officials, and community members who resent increased state involvement in curriculum.

Scheduling Challenges:
- Current calendar diversity would add difficulty to TCA scheduling.

While many schools operate on a late August to early June calendar, there are some schools that operate year-round, as well as schools that begin classes early in August and end school toward the middle of May. As a result, it will be difficult to design a TCA testing calendar when students in some schools would only be prepared to take unit assessments long after or before the majority of students in the schools with traditional calendars. Preserving the security of the item bank is a specific concern in any situation in which some students are exposed to exam items before others.

CONCLUSION:

There are advantages and limitations to a TCA model. However, if designed carefully and implemented with education and outreach for all involved, the model has the promise to produce higher-quality information on student achievement as well as to help assess their understanding of knowledge and skills that are critical to their success. With richer information on student mastery, teachers will be better able to meet the unique needs of their students.

WORKS CITED

By Jennifer Preston and J. Eric Moore

The Financial and Business Services Area is in its fourth year of the Research Intern Program. The Program is designed to help build a quality research program within DPI to supplement and supply data for discussions related to procedural, process, and policy changes. This year’s program included students from the Duke University master’s program in Public Policy, a graduate of the University of North Carolina at Chapel Hill master’s program in Public Administration, and a doctoral student from North Carolina State University in Public Administration. The intern program is managed by Christi Chadwick (919-807-4029) and Kayla Siler (919-807-3824) or intern_research@dpi.state.nc.us

NC DEPARTMENT OF PUBLIC INSTRUCTION :: June St. Clair Atkinson, Ed.D., State Superintendent :: 301 N. Wilmington Street :: Raleigh, NC 27691-3825
In compliance with federal law, NC Public Schools administers all state-operated educational programs, employment activities and admissions without discrimination because of race, religion, national or ethnic origin, color, age, military service, disability, or gender, except where exemption is appropriate and allowed by law.

Inquiries or complaints regarding discrimination issues should be directed to: Dr. Rebecca Garland, Chief Academic Officer
Academic Services and Instructional Support :: 6388 Mail Service Center, Raleigh, NC 27699-6388 :: Telephone: (919) 807-3200 :: Fax: (919) 807-4065

Works Cited