DATA RECOGNITION
DRC
CORPORATION

Study on the Feasibility and Cost of Converting the State Assessment Program to a Computer-Based or Computer-Adaptive Format

F I N A L R E P O R T

EXECUTIVE SUMMARY

June 25, 2007

EXECUTIVE SUMMARY REPORT

Data Recognition Corporation (DRC) is pleased to submit the Final Report on the Feasibility and Cost of Converting the State Assessment Program to a Computer-Based or Computer-Adaptive Format (CBT/CAT).

Overview

This report focuses on the 14 components outlined in the South Carolina Request for Proposal and brings together the expertise of DRC's research, assessment, policy, and psychometric staff, as well as that of the following independent consultants: Dr. Steven Wise, Institute for Computer-Based Assessment at James Madison University; Dr. Richard Luecht, Center for Educational Research and Evaluation at the University of North Carolina at Greensboro; and Dr. John Poggio, Center for Educational Testing and Evaluation at the University of Kansas.

GOALS

The goal of the report is to provide SC policymakers and stakeholders with an objective picture of relevant research, national and state-specific information related to assessment, clearly delineated advantages and disadvantages, and projected costs for implementing a computer-based or computer-adaptive assessment program to inform South Carolina's decisions regarding the future of the state's high-stakes testing program.

Additionally, the South Carolina Team (Tammy Mainwaring, Project Manager and Change Manger, CIO; Dr. Paul Horne, Director of Curriculum and Program Review, Education Oversight Committee; Elizabeth Jones, Education Associate, Office of Assessment, State Department of Education; and Deidre Appleby, Education Associate, Office of Technology, State Department of Education) communicated five driving factors to be considered throughout this report:

1. The desire to receive test results back more quickly;
2. The need to address the concerns regarding the amount of time spent on testing;
3. The desire to get increased diagnostic information;
4. The desire to identify costs associated with computerized testing; and
5. The desire to have the best quality assessment program possible.

These factors have implicitly or explicitly informed the information and conclusions provided for each component.

ASSUMPTIONS AND CONCLUSIONS

As will become apparent throughout the report, a transition to computerized testing, in whatever form, would require choices and trade-offs, not only in terms of funding technology infrastructure, but in terms of training, testing windows, psychometrics (how measurement data are collected and used), the inclusion of constructed-response items, reporting turnarounds and diagnostic information.

There is no "silver bullet" or one right answer.

With that in mind, the report also includes a number of recommendations (as required by the RFP). These recommendations are based on the experiences of other states that have moved to computerized testing, extensive research, No Child Left Behind (NCLB) requirements, and the results of school and district surveys conducted within South Carolina.

Please note that since much of this report focuses on the feasibility and costs for converting the state assessment program to a CBT/CAT delivery mode, no assumptions were made as to *which* test items would be delivered via computer (i.e., existing South Carolina items; new, custom-developed items; existing item banks or tests offered by various vendors). Therefore, no cost estimates are given for actual test item development, for test item adoption, or for the conversion of existing SC items from paper/pencil to a computer delivery mode. These costs are outside the scope of this study and will be addressed by state policy decisions.

The analyses of various computerized test delivery systems provided in this report are confined to a sampling of those systems and vendors that are currently offering high-stakes computerized summative assessments in the United States. Many districts in South Carolina administer computerized formative assessments. The recommendations outlined in this report fully support the use of formative assessments, and they are included in Components 13 & 14 as desirable companions to summative tests. However, the system capabilities of various formative assessments were not addressed in this report.

Cost estimates and recommendations, instead, focus on infrastructure needs and address only computerized *delivery* systems used for high-stakes, summative testing and projected delivery system and transitioning costs.

Likewise, the recommendation for a formative assessment component makes no assumptions as to whether the state may develop or adopt a formative assessment item bank, or leave formative assessment system choices to districts using the state-approved list of formative assessments. Rather, the recommendations focus on ensuring there is a sufficiently large formative item bank to provide diagnostic information, and that those formative items are both aligned to state standards and similar in structure, format, and delivery to the high-stakes, summative test items.

Lastly, with regard to the projected costs for infrastructure improvements needed for a transition to computerized testing (Components 7 & 9), such investments must be viewed within the broader context of improved instruction and work-force readiness for South Carolina students. The same technology infrastructure used for testing should be used in daily instruction throughout the year. The integration of technology in instruction is a goal in and of itself. The assessment mode should mirror instruction, and vice versa. For example, if students are composing, editing, and publishing their writing via computer, to test students on writing skills via paper/pencil would not mirror instruction or their day-to-day reliance on technology in the classroom.

As such, these investments should be seen within an overall context of an improved and equitable instructional environment for all students.

Electronic Testing: Terms and Introductory Concepts

Computers are revolutionizing education in general and testing specifically. The Feasibility Study focuses on test administration, although many related issues, including test design, content, accommodations, and reporting, must also be carefully considered. To avoid confusion in an emerging field with many similar-sounding terms, the study begins with the definition of key terms. The definitions apply to this document only. Note that other authors and other documents, including those listed in the references, may make different distinctions.

Electronic testing (eTesting) refers to any assessment presented to the student via a computer screen and with which the student interacts via a keyboard, mouse, or other pointing device. It may or may not involve the Internet and World Wide Web. It may or may not involve branching decisions to tailor the test to the student. *Computerized testing* is considered to be synonymous with eTesting and will be used interchangeably.

Computerized adaptive testing (CAT) refers to an electronic test that uses branching to tailor the test to the student's level of proficiency. In general, no two students will receive the same set of items. This topic will be discussed in detail in the section of this report entitled: *Computerized Adaptive Testing*.

Computer-based testing (CBT) will be restricted here to refer to the administration of fixed forms, in contrast to CAT. Generally, all students will take the same items. However, it does include the case of multiple versions of the form with the same items in different orders (i.e., scrambled forms). It could also include several parallel fixed forms. *Parallel fixed forms* means all forms are built prior to testing using the same content and statistical specifications but each form will contain different items. Both scrambled forms and parallel forms are used to enhance security.

Component 1 – Description of State Assessment Programs which are Computer-based or Computer-adaptive

STATE OF THE STATES CONDUCTING CBT/CAT

In February 2007, a questionnaire was e-mailed to all state departments of education inquiring about various aspects of state initiatives in computerized testing. All fifty state departments responded.

Eighteen states indicated they offer some level of statewide, summative testing via computer. Of those, six states offer across-the-board computerized testing in grades 3-8; however, only Idaho, Oregon, and Wyoming make their computerized NCLB tests mandatory for most students.

Representative comments from states using CBT/CAT include:

> **Idaho:** The state "has been happy with our computer-delivered tests. Change is always hard, and some school districts were very apprehensive when we stared the computer-delivered tests. At this point however, there would not be one school district that would wish to go back to paper/pencil."

> **Wyoming:** "…Students and schools really like the online testing, but it does impact the use of computers and tech resources during the testing window. Wyoming districts are tech ready and early adopters. They were willing to experience the glitches in order to have more rapid reporting and better control of test security."

Most states that are not currently providing computerized testing indicated a strong interest in moving into that arena, and states that are providing limited computerized testing indicated plans for expansion of their programs.

Of the states providing insights into why computerized testing is not being offered or is being offered only on a limited basis, the following challenges were cited: **capacity**, **connectivity**, **funding**, **technical support**, and **the need to ensure a "full-service" vendor for both paper/pencil and computerized delivery methods**.

Representative comments from these states include:

Vermont: "We are very interested in computer-delivered testing, but our issue is capacity. We still have many schools with dial-up connections, and insufficient terminals to make web-based testing feasible."

Massachusetts indicates that the state is considering options for online testing in 2008; however, state funding is not expected to cover the requisite costs for this initiative. Thus, "online testing plans are now on hold."

Indiana (which offers high school end-of-course tests online): "Special challenges include finding a vendor that is a one-stop shop for delivering, scoring, reporting, and psychometric services for CBT and P/P assessments (simultaneously)."

STATE IMPLEMENTATION STRATEGIES FOR CBT/CAT

Most states that are moving their testing programs into the online arena have focused or will initially focus their computerized testing initiatives on non-NCLB grades/subjects (e.g., end-of-course tests, writing, geography), as they are less "high stakes" for NCLB and/or state-level accountability.

COMPUTERIZED TESTING FOR STUDENTS WITH DISABILITIES AND ENGLISH LANGUAGE LEARNERS

At least two states, Kentucky and Minnesota, have initially implemented computerized testing for students in need of special testing accommodations, since computer-delivered tests can offer increased access, standardization of administration, and flexibility in the form of text-to-speech software (as opposed to human readers), adjustable fonts (style, size, color), translators, streaming video, and other features.

OVERVIEW OF STATE-LEVEL STANDARDS FOR STUDENT KEYBOARDING AND OTHER TECHNOLOGY SKILLS

This study has also reviewed how various states ensure that students develop technology skills, such as keyboarding, that would be required for computerized testing. While 42 states have technology standards for students in place, relatively few states have mandated "stand-alone" technology skills standards that specify the skills to be mastered at each grade or grade cluster, as opposed to technology standards embedded in content-area standards. South Carolina has embedded standards.

Of the states that have "stand-alone" standards, many require teaching keyboarding and word-processing skills, as well as multimedia, spreadsheet, and database skills, beginning at the K-2 grade levels. South Carolina may wish to develop or adopt such standards to ensure that students, at an early age, are being prepared for the demands of the 21st century workplace, as well as computerized testing.

Component 2 – A review of the Literature on the Comparability of Scores Obtained by Examinees when Assessments are Administered by Computer Rather Than Paper and Pencil

When moving from paper/pencil based to computer-based testing (CBT) in today's environment, it is essential that the transition be seamless. There must be no artificial gains or losses associated with the change. NCLB requirements mandate that any such transition incorporate all appropriate comparability studies to establish the consistency of the measures.

Comparability studies are validity studies. The current paper/pencil is the baseline that has established the construct, the scale, and the performance standards. Paper/pencil currently defines valid. For CBT scores to be comparable to paper/pencil, the scores must be interchangeable. There can be no change in item difficulties associated with the mode of presentation.

If scores from different modes are not interchangeable, it may be possible to make them comparable by shifting the electronic testing scores by the effect size. However, because different items and item types may be affected differently, more elaborate approaches such as multi-trait, multi-method techniques may prove more appropriate and effective than simple mean shift equating.

There appears to be a small advantage (0.02 mean effect size) for students taking a multiple-choice test on computer. This is based on a meta-analysis of 79 studies, of which 21 reported higher scores for the CBT and 8 reported higher scores for paper/pencil. The data suggest differences among content areas and grade level, with most negative results for CBT associated with mathematics and elementary school students. However, sample sizes are small when broken down to this level. The studies dealt with K–12 assessments and were conducted since 1997.

There are very few studies looking at direct measures of writing and those that do present inconsistent results. Of the three large studies of this issue, Kingston (2004) reported a moderate advantage for students taking a direct writing measure on computer; Sandene et al. (2005)

reported no difference; Pearson Educational Measurement (2006) reported an advantage to students taking the test on paper.

There appears to be no disadvantage to students based on their gender, race, ethnicity, or socio-economic status. While there is a technology gap among subgroups, there are no consistent data indicating a disadvantage of CBT because of it. Additional local studies would be advised.

Regardless of their relative performance on paper/pencil and CBT and previous computer experience, students expressed a strong preference for CBT. The students' level of satisfaction could be quickly diminished if they are not permitted to review and change earlier items. Systems that permit item review and answer changes received much higher levels of student satisfaction.

The evidence regarding the impact of computer familiarity on performance on an electronic test is inconsistent. It may be minimized by giving students sufficient experience and realistic practice tests with the computer administration system before the test is administered. Teaching all students keyboarding skills may need to be included in the curriculum.

It is perhaps obvious that, for any type of testing with any mode of delivery, the assessment must match the instruction. When instruction is via a computer, students perform better when tested via a computer. When a computer is not integral to instruction, electronic testing typically results in lower scores. This may compel the implementation and requirement of content standards for technology-related curriculum.

Component 3 – A Determination as to Whether the Conversion of the State Assessment Program to a Computer-based or Computer-Adaptive Format Will Satisfy the Federal No Child Left Behind (NCLB) Requirements

CURRENT NCLB REQUIREMENTS

No requirement in the current NCLB legislation specifically precludes computer-based testing or computer-adaptive testing, as long as the computerized testing meets all NCLB requirements that pertain to all tests, whether they are administered via computers or paper/pencil (e.g., tests are aligned with state standards and on grade level). However, there are NCLB requirements that are particularly relevant to states transitioning to computerized testing. These include requirements for comparability among modes of administration (i.e., no mode-related construct-irrelevant variance), assurances of score comparability for Adequate Yearly Progress (AYP) reporting consistency via equating or otherwise, and on-grade-level assessment items for AYP reporting purposes.

COMPARABILITY

NCLB requires documented evidence of the comparability of the computer-based or computer-adaptive test administration with previous or concurrent paper/pencil test administrations. Establishing comparability would be important even if the state were to transition entirely to computerized testing in a single year, since the previous score data, and associated information, reported both within the state and to the U.S. Department of Education, would have been from paper/pencil administered tests. Meeting these comparability requirements is not insurmountable, as many states have done so successfully through careful planning and using psychometrically sound methodologies.

CONSTRUCT-IRRELEVANT VARIANCE (TESTING MODE EFFECT)

Secondly, a transition to computerized testing must ensure that the test administration method does not interfere with a student's test performance (i.e., construct-irrelevant variance). It is also important to ensure that the mode of administration itself (i.e., the computer) does not introduce unintended and unforeseen variables into the testing situation, as the intent of a

computerized test is to measure students' educational achievement and not to assess their computer expertise. Results from comparability studies should help to address the question of whether students who have had less computer experience are impacted differently by the electronic test administration.

With careful planning and implementation, these concerns can be addressed to satisfy NCLB requirements.

SPECIFIC ISSUES RELATED TO COMPUTER-ADAPTIVE TESTING AND NCLB

Lastly, the USDOE has determined that all test items used for AYP determinations must be on grade level and measure a component of the state's content standards for that grade. In the traditional CAT environment, the computer algorithm chooses the "best" items to determine a student's performance level on a particular skill, even if those items fall outside the assigned grade-level for that skill. That is, the CAT algorithm selects and administers the item closest to the student's estimated level of achievement regardless of the grade nominally associated with the item. Therefore, a traditional CAT system would not meet NCLB requirements.

However, two states have implemented variations of traditional computer-adaptive testing that still meet the on-grade-level requirements of NCLB: Idaho and Oregon. Idaho uses a core of grade-level items for reporting purposes, and then adds potentially off-grade level CAT items for further diagnostic information. Oregon uses a CAT system, yet all items *within* the system are on-grade-level.

Component 4 – Recommendations Regarding Subject Area Assessments to be Computer-based or Computer-adaptive to include a Recommendation Regarding Order of Implementation

An informal questionnaire was distributed as the South Carolina Educators for the Practical Use of Research (SCEPUR) Annual Conference and the South Carolina Middle School Association (MSA) Annual Conference. This questionnaire was intended to elicit the opinions and impressions of front line educators about the advantages and disadvantages of electronic testing. Overall there was strong support for electronic testing, assuming it includes rapid return of diagnostic reports to the teachers.

The educators responding to the questionnaire indicated a preference that electronic testing begins with reading and mathematics because of the central role these content areas have in the curriculum. This must be interpreted in light of the expectation for early reports. Because of the importance of reading and mathematics, these are the areas for which the teachers are most anxious to get more data faster.

There was a concern from the educators that it may be unfair to assess writing on computers in elementary grades because of the lack of practice composing on computers. The Web survey, sent to all schools and districts, indicated widespread practice with computers including writing, which began early and continued through grade 12. The percent of students using computers reached it maximum by grade three but the total hours spent using computers continued to increase until high school.

The Expert Panel strongly advised to begin electronic testing small, using volunteer participants to build capacity, confidence, and support. It is critical that the early forays succeed and that back-up plans be included from the beginning.

There are many ways one might begin to implement electronic testing and any approach will involve trade-offs. Many of the considerations relate to policy rather than technology or psychometrics. One can argue for beginning with a non-NCLB content area for the early stages. This would relieve some of the pressure until comparability studies can be done and the

infrastructure thoroughly tested. On the other hand, teachers in their responses to the questionnaire favored the two central NCLB areas of reading and mathematics.

Component 5 – Feasibility of Inclusion of Constructed-response Items as Part of the Writing Assessment and Feasibility of Movement of the Writing Assessment to a Separate Administration

Like electronic testing in general, automated essay scoring (AES) is also probably inevitable in some form and role. The existing models are promising, functioning with reliabilities similar to human scoring. There is also the potential for faster reporting of results and enhanced diagnostics. This may make it more appropriate and attractive for formative testing and instruction than for large-scale assessment and accountability systems.

For the immediate future, any use of automated scoring in high stakes situations will almost certainly be run in conjunction with human scoring. Most systems for AES will continue to require some amount of human scoring for calibration or security. The parallel systems with one human score and one computer-generated score are appropriate both to maintain public confidence and to safeguard against unforeseen anomalies the computer algorithms were not prepared to handle.

In terms of AES, models for scoring writing are the most developed and the most used. Short-answer constructed responses (CR) present greater challenges to the computer scoring engines. While it may seem counter intuitive that short responses are harder to score than long responses, it is actually an issue with the number of words written. A response with very few words does not adequately sample the student's proficiency to permit a computer algorithm to make an inference. The regression, artificial intelligence, and latent semantic analysis models require more information to function correctly.

Very short responses are perhaps best scored with pattern matching. These are quickly evolving beyond exact matches to more complex natural language processing and fuzzy logic applications. This now includes algorithms that can identify the appropriate information in a reading passage needed to answer and then evaluate whether the student's response agrees well enough to receive credit. The combination of constructing the key and evaluating the student's response make this approach very attractive for formative assessment, but perhaps not for large scale, high stakes.

Another approach to earlier reporting of writing assessments is earlier testing. Using a separate window for testing writing allows more time for hand-scoring and makes it possible to return preliminary writing results before the end of the school year. This poses no significant psychometric issues but does present two logistical problems. A separate writing window means another test administration for the schools, with all the associated shipping, receiving, rescheduling, and disruption that implies. It also poses an additional problem for the contractor if the scores from the two (or more) testing windows must be collated for reporting or analysis.

An analysis of the reliability, precision, and decision consistency associated with CR indicates there is no psychometric argument for or against CR. In terms of the amount of information gained for the time and cost, it is more efficient to use MC items. However, the validity of the assessment, in particular writing assessment, is enhanced by the direct assessment of the skills intended through constructed response and extended response items.

Component 6 – Hardware, Software, Staffing, and Training Requirements at the State, District, and School Levels to Administer Statewide Computer-based or Computer adaptive Assessments

Although more than 177,000 computers are available to students within the state, 43% of those computers would not be available for testing. Among the computers available for testing and the technical specifications of the testing products analyzed, the greatest barrier to testing (the system requirement with the highest "failure" percentage) is processor speed; 36% of the computers would need a processor upgrade, and 14% of the computers could only support one of the testing products analyzed. The second major barrier is system memory; 13% of the computers need an upgrade to reach 256 MB RAM system requirement. A majority of the computers currently run a supported operating system (91%) and supported browsers (88%). Based upon the average age of the computers, many of the computers that would require processor or memory upgrades may already be scheduled for replacement.

Most of the schools utilize content filtering and firewalls. These products will still allow testing, but would need specific configuration changes to authorize certain Internet addresses during testing. Other system configurations may need to be adjusted during testing. This may include changing pop-up blocking, allowing session cookies, or enabling or disabling applications that run automatically (e.g., email notifications or virus scanning software).

A key area within the technology infrastructure is how many students will be able to test simultaneously given the current bandwidth at the school. Although the bandwidth needs may be reduced by utilizing a cache for test delivery, the number of students supported by each Internet connection may not be adequate to support the levels of concurrent users anticipated. The data gathered within the survey does not include information about whether multiple buildings share Internet connections, or if there are dedicated lines per school. This will have a great impact on performance during testing. Approximately 24% of the computers available for testing could support 120 students testing simultaneously per T1 line, assuming little or no other online activity during assessment time.

Three of the vendors analyzed have products that include multimedia test delivery. This includes, at minimum, a text-to-speech feature that may be utilized for students who require a read-aloud accommodation. These products may require a more robust system configuration or the implementation of a caching server to deliver the test content to all students within acceptable performance levels.

The staff members who responded to the survey provided general comments regarding electronic testing. Based on these comments, staff members are concerned with the number of computers available for testing in order to complete tests for all students within the testing window, as well as the availability of technical staff to handle computer glitches during testing. Establishing support from key stakeholders throughout the state, including state representatives, technical staff, and test coordinators, is possible through frequent communication and a thorough training program. Such a training program should include demonstrations, simulations, and clearly defined processes for handling any alerts that occur during testing. Aside from network and system configuration, it seems likely that staff members with basic computer skills would be able to handle most of the alerts that occur during testing. Fostering partnerships between the state and testing vendor, as well as between technical staff and test coordinators, is key to increasing the comfort level of staff regarding electronic testing.

Components 7 & 9 – Costs to the State of Converting the State Assessment Program to a Computer-based or Computer-adaptive Format; Costs to Bring the State, Districts, and Schools to Needed Capacity for Delivery and Maintenance of a Computer-based or Computer-adaptive State Assessment

When any state is considering a transition to electronically delivered assessments, cost is a key factor. The desire for technology advancement and providing the best and most useful testing environment for students must be balanced with budgetary considerations. A common misperception is that the reduction in printing, packaging, and distribution costs associated with a paper-based assessment program will lead to immediate and dramatic cost reductions. However, the experiences of many states delivering large-scale assessments electronically have shown this not to be the case. The introduction of new costs such as:

- Infrastructure acquisitions and upgrades necessary to bring schools to desired levels
- Increased item development costs due to reformatting of existing items and/or expansion of the item bank
- Changes to necessary staffing and training plans
- Adjustments to the configuration of computer labs to accommodate large-scale assessment
- Additional comparability studies necessary to meet NCLB guidelines

All of the states surveyed as part of this study showed an initial upward spike in costs during the first years of implementation.

For South Carolina, the largest single cost or investment factor to consider is placing enough computers and supporting equipment (servers, networking, and bandwidth) into the schools to allow them to deliver the planned summative assessments inside of a fixed testing window. Contained in this study is an analysis of the overall student to computer ratio across the state along with a more detailed analysis of the students to "testing computer" ratio, which focuses on the number of computers that could be made available for testing at any given period

of time. This study also used an analysis of students to computer ratio of other states delivering a high percentage of their summative testing programs online to establish a baseline or target ratio. The analysis the current number of testing computer available and the established target is show in the following table:

Total Computers Needed to Accommodate Electronic Testing

Total computers needed to achieve a 4:1 student-to-testing computer ratio	162,500
Total computer available for testing (based on survey data)	100,372
Number of additional computers needed	**62,128**

This study recommends making up the deficiency in the number of computers through a combination of the acquisition of new computers on the open market and by making adjustments in computer location to allow a greater percentage of the computer current in schools to be made available on testing day. Using this method, along with an analysis of the current market price of hardware, we estimated the following infrastructure investment necessary to prepare the state for full-scale implementation:

Summary of Infrastructure Cost

Cost Item	Amount
Acquisition of additional computers	$32,500,000
Acquisition of additional hardware (servers, networking equipment, cabling, power)	$9,750,000
Upgrades to existing computers (memory, processor, operating system) to meet recommended system requirements for test delivery	$12,000,000
Total recommended infrastructure investment	**$54,250,000**

The total cost of implementation for any state depends on a multitude of factors making it difficult to provide total cost down to a single number. This presents both a challenge and an opportunity for budget driven stakeholders. While there is a challenge in attempting to develop a budgetary number for appropriation purposes, the state can have a large amount of flexibility in tailoring the aggressiveness of an implementation plan to the funding available. More aggressive plans would come with a higher degree of up front costs, but faster progress toward cost savings. A less aggressive approach would allow the technology investment to be spread out over a longer period of time, but would also come with a longer return-on-investment period.

In the State of South Carolina, cost is clearly a major concern for staff in the districts and schools. Virtually every testing site felt that they would need additional funding for technology and staff to be able to accommodate a transition from the current delivery model to an electronic one. To counteract this perception, the state will need to establish clear policies and communication plans with test coordinators and administrators to fully understand and address the needs of individual sites. The state will need to assist districts that have not made technology investments to keep pace with the times.

Component 8 – Current State, District, and School Capacity, to Include Personnel, for Administering a Computer-based or Computer-adaptive State Assessment

This section of the study focused specifically on profiling the personnel needed to implement and administer computerized testing both at the district and at the school level. The profile consisted of the number of information technology (IT) staff currently available, the average number of IT staff available for each school type, the levels of experience with computerized testing, and an estimated number of staff needed if South Carolina implemented computerized testing. Component 8 also considered the types of resources required to address the complexity of implementation, the link to instruction, and the importance of a strong Assessment, Technology, and Education Policy Partnership.

RESOURCES REQUIRED TO ADDRESS COMPLEXITY OF IMPLEMENTATION

Implementation of online tests can be complex because both the technology and the assessment elements of the testing process have to be aligned and working well.

Personnel in charge of supporting the online testing environment will need to have the appropriate skills to address the following challenges:

- Addressing hardware and software requirements delineated in Component 6 (memory RAM, processing speed, bandwidth, operating systems);
- Preparing local infrastructure for testing readiness (system requirements, needed computer upgrades, and sufficient testing of local software so that it will not interfere with the testing system);
- Facilitating the administration of the online test and monitoring the testing lab or classroom environment to ensure computer readiness, test security, and test administration issues. For example, test-day issues may include system re-boots, error messages, or unforeseen computer glitches); and
- Monitoring the online test administration to ensure that appropriate testing procedures are followed and that test security is maintained.

LINK TO INSTRUCTION

An important element in the successful implementation of computerized testing is the alignment of assessment and instruction. The assessment venue should reflect the mode of delivery of instruction. Instruction solely in one medium while assessing in another should be minimized or avoided so that student performance on the assessment accurately reflects student learning.

ASSESSMENT, TECHNOLOGY & EDUCATION PARTNERSHIP

Critical to the success of administering online assessments is the partnership among the assessment, technology, and education representatives within the state. Each group plays an important role in developing and implementing successful, user-friendly, and instructionally relevant online testing. Each group offers and brings a different perspective to the table, and a close relationship among these groups is absolutely essential for the success of any online testing initiative.

FINDINGS

South Carolina is developing a statewide infrastructure of state, district, and school personnel who have expertise in implementing and maintaining robust technology systems that can support both computerized testing and well-grounded computer-enhanced instruction. This infrastructure should continue to grow incrementally in numbers and mature in expertise as a computerized testing program is implemented. In this manner, the state can support the integration of assessment and instruction so that the resources allocated to computerized testing can also serve to support computer-enhanced instruction.

The specifics of any operational planning for personnel to support a conversion to computerized testing depend in large part on the specific factors and requirements of the plan chosen (e.g., specific testing system chosen, aggressiveness of the implementation plan).

However, the following summary points are worth noting as implementation plans are considered:

- There is a wide range of experience levels among technical and testing staff, but the vast majority had some level of experience with computerized testing (one or more instances). District test coordinators and district technology coordinators tend to have the most experience.
- Survey responses indicate that South Carolina has a pool of district-level individuals with computerized testing experienced to draw from for future actions planning, training, etc. (i.e., 70% or more of district technology coordinators or district test coordinators have had five or more experiences with computerized testing). The level of experience is lower at the school level, but 50% or more of school technology coordinators, school test coordinators, and school administrators have had three or more computerized testing experiences.
- For computerized testing, personnel needs for testing setup and administration are likely to shift slightly. Responses indicate that computerized testing would require somewhat fewer temporary personnel, and fewer teachers would need to be trained in testing. However, respondents estimated that more technical coordinators and test administrators would be required at the school level.
- Responses show that district-based technicians are available to most schools. However, it is likely that district-based technicians would provide service to several different schools. Therefore, if a decision is made to implement computerized testing, South Carolina may need to take into account the number of schools that technicians are expected to support.

Component 10 – Issues Regarding Window of Administration, Test Security, and Need of a Backup System for State, District, School, and Classroom Purposes

This section discusses the security implications of moving from a paper/pencil-delivered assessment program to an electronically delivered program. The security of an assessment program can be large and multi-faceted. When considering a move to electronic delivery of an assessment program, there are several factors to consider:

- Item security and exposure
- Security of the test delivery system
- Security of the student registration system
- Security plans and audits
- Backup and failover plans and strategies to ensure system reliability
- Placement of computers in labs

States have an obligation to work in close conjunction with their testing vendor(s) to develop comprehensive security and failover plans that addresses all major risk factors and ensure a safe and successful delivery of the program. In order to maximize the overall security of the assessment program, this study recommends the following courses of action:

- The state should work closely with its testing vendors to define the overall security requirements for the testing program. These requirements establish the baseline for testing and assessment.
- The state should take a "trust but verify" stance with assessment vendors. Careful review and approval of vendor security plans are a must. The state may also look for vendors to meet nationally accepted security certifications. If necessary, the state may request an independent security audit of the assessment vendors.

- The state should work closely with vendors to establish backup and failover plans for as many situations as possible. These plans may vary in the cost and complexity in relation to the potential risk. However, for each plan, a clear communication approach should be developed and ready to implement.
- The state should establish communication channels that allow test administrators and technology coordinators to easily communicate with one another. By facilitating communication between parties in the field, the state can address issues more quickly, share lessons learned, and reduce discomfort levels when issues do arise.

Component 11 – Delivery of Results for Schools and Students and the Ability to Provide Instructionally Informative Results to Districts, Schools, Teachers, and Parents

As South Carolina considers moving toward more timely, instructionally informative reports, several aspects of other state testing programs are reviewed to provide context for South Carolina's decision-making process. This information includes testing windows, report turnaround times, testing resources, and whether the testing programs are paper/pencil-based, CBT/CAT, or some combination. This section concludes with a discussion of options for providing Computerized Testing Score Reports that are linked to instruction.

SUMMARY OF FINDINGS FROM OTHER STATES

Overview of Test Designs

Thirty-three states offer multiple-choice tests with a constructed-response component, excluding writing as a separate test. Thirty-five states, including South Carolina, assess writing in at least one grade. While the scoring of open-ended and writing assessments generally increases the time between testing and reporting, most states have determined that an assessment that goes beyond multiple-choice is worth the cost and the slower reporting.

Reports

Most states (46) provide some sort of reporting at a strand/standard level, usually by raw scores or percent correct. The minimum number of points used for reporting at the standard level range from 1–12, with a minimum of 6–8 being the most common. South Carolina is one of a handful of states that reports scores only at the overall content/subject level.

For any number of items, it is preferable to report in the scale score metric rather than raw scores. This removes the effect of specific selection of items and facilitates comparisons across years and across standards.

Testing Windows

There is a great variety of testing windows among the states. Many states have a single day for testing each subject, while others have testing windows of a month or more. Testing windows are often longer for CBT/CAT than paper/pencil due to limited computer access. However, many paper/pencil-based testing programs also have extended testing windows.

The time of year testing is conducted also varies—at least seven states test in the fall rather than the spring. At least nine of the 35 states that assess writing separately use a writing window earlier in the year to allow additional time for hand scoring, with an additional five states assessing English language arts/reading/writing in some combination earlier in the spring.

Report Turnarounds

The time between testing and reporting varies widely by state and testing program. For paper/pencil tests, the turnaround time generally ranges from 6 to 16 weeks. Computerized testing tends to have a faster turnaround. Preliminary scores can be immediate (e.g., Idaho, Oregon—multiple choice only), or, Virginia, which does post-equating of certain CBT forms, has reporting targets of less than two weeks. Other computerized tests do not provide immediate reporting: Florida (MC only, six weeks); Texas (two weeks); Mississippi (pass/fail rosters in three weeks); Wyoming (MC, second week of testing window; short CR available four weeks after close of testing window); and West Virginia (writing, reported in 60 days).

Additional Testing Information Provided by States (e.g., Sample Items, Released Tests)

Most states provide sample items or released test forms for every grade and subject, every other year. Four states provide sample tests; 17 states provide practice tests; 6 states provide released tests; 26 states provide released items; and 19 states provide sample items.

State Score Reporting That Provides Activities and/or Instructional Links for Parents and/or Teachers

A few states are providing summative testing results that link assessment with instructional activities for parents and teachers. For example, Ohio's score reports provide a "Next Steps" section, which includes activities that parents can do with their children, based on their child's performance on each standard.

The Grow Network/McGraw-Hill's *MyGuide Program* links student-specific results to activities via "personalized study guides" for high school retesters in Texas, Arizona, and California. Students who have not passed all of their state's high school exit exam may visit a state-specific website, enter their scores, and a customized study guide is generated, including guided practice and tutorials. To assist students in using their "personalized study guides, resources are also provided for teachers, tutors, and parents. Finally, the Grow Network, through hard copy student reports and an associated website, is providing parent links to home activities for several states.

Component 12 – Advantages and Disadvantages of Converting the State Assessment Program to a Computer-based or Computer-adaptive Assessment

This component of the study presents an analysis of the advantages and disadvantages of computerized testing, focusing on both CAT and CBT. A panel of experts and researchers representing states that have implemented CAT or CBT was convened to provide South Carolina with the latest, most accurate information about other states' initiatives in computerized testing. Additional sources of information included a survey of all 50 states and reviews of current pertinent literature.

Advice offered by Expert Panel members included the following points:

- Set the expectations for and the goals to be achieved through CBT or CAT.
- Ensure that districts and schools have the necessary technology, bandwidth, hardware, and training to be successful in implementing the program.
- Make certain that the Request for Proposals is very specific and details the minimum threshold of PC requirements.
- Lay out an implementation plan that has incorporated input from all stakeholders.
- Make certain that the legislature is prepared to fund the up-front costs.
- Select a knowledgeable committee to serve as ongoing advisors for the computerized testing initiative.
- Call for an independent third-party assessment of the chosen vendor's security and system.
- Build the assessment specifically for online delivery.
- Use scientifically-based research and psychometric "best practice" to guide decisions.
- Plan for sufficient technology support.
- Ensure that training for districts is a team effort between the state and the vendor.
- Prepare for resistance.

- Elicit advice from other states that have implemented computerized testing programs.
- Start with a practice or pilot test before implementing high-stakes computerized assessments.
- Link formative or practice tests to the online summative assessments.
- Inform policymakers and stakeholders that initial costs will probably be higher than with a completely paper-based assessment system.
- Foster the idea that the way students are tested should always reflect the way they are taught, i.e., the use of technology should be an integral part of classroom instruction.

COMPUTER-BASED TESTS/TESTING: ADVANTAGES AND DISADVANTAGES

Immediacy of Results

Among the most notable features associated with computerized testing is the potential for immediate or almost immediate test results. To obtain immediate results, however, requires that the assessments is either comprised entirely of multiple-choice with short-answer test items that can be scored objectively using artificial intelligence systems.

As described in Component 11, many states have chosen not to abandon the assessment of written composition in order to expedite the reporting of results but instead administer their writing assessments earlier in the school year, either through a separate writing test or administering their entire English language arts test earlier.

Providing immediate test results also typically requires that the psychometric methodology for equating tests must be changed from a more robust "post-equating" design, which South Carolina currently uses, to a "pre-equating" design, which is less optimal. Continued use of a post-equating model does not preclude the possibility of computerized testing, but it does preclude the possibility of immediate results. Test results would not be available immediately following testing without a shift to a pre-equated model. The state would need to carefully consider this shift from post-equated to pre-equated models and discuss this with its Technical Advisory Committee to evaluate how such a shift would impact the transition from paper-based to computerized testing and/or the co-existence of both paper-based and computerized tests for a period of time.

If constructed response or extended response items remain in the South Carolina assessment program and human scoring continues to be used, a compromise may be the use of computerized testing with minimally delayed reporting. Using computers to administer the entire test, the constructed responses and/or extended response questions would automatically be routed to a scoring center where they would be scored by human readers. This quick transfer of student responses for human scoring would reduce turnaround time for scoring extended and constructed response items but would prohibit the immediate return of total test results.

Cost Efficiency

Moving the South Carolina assessment program from a system that is paper-based to one that is delivered almost exclusively via computer will likely not result in a significant reduction in overall assessment costs, particularly in the early years of the computerized assessment program. Even though the production of test booklets and answer documents and the shipment of large quantities of materials will be greatly reduced, and eventually almost eliminated, certain components of any high-stakes assessment program must remain, irrespective of the mode of delivery. Specifically these are: development of test specifications; item development and review; field-testing; test security procedures; test form development; human scoring of CRs/ERs (should these be retained); psychometric analyses to ensure proper equating and standard-setting; and creation of paper score reports for distribution to districts, schools, and parents.

Additionally a transition to computerized testing would require test comparability studies, as required by NCLB. Finally, the initial investments in hardware, software, capacity, and staffing will obscure cost savings in the early years of computerized test administration, but may be offset by enhancing access to technology in the classroom and allowing technology to become a more integral part of classroom instruction.

Instructional Uses

Once schools have the necessary expertise and capacity for computerized testing, the infrastructure (hardware, software, connectivity, networking, and technology expertise in all districts) can be utilized on a daily basis to support and extend instruction. Teachers will be able to administer formative assessments and use the results to inform their instructional practices,

monitor their students learning progress through an online database, and create innovative lesson plans that maximize the use of technology.

Usefulness and Accuracy of Student Data

The speed with which data are available to educators allows rapid analysis and evaluation of performance. Computerized tests allow educators to review results and make instructional and evaluative decisions on behalf of students while the information is current. When tests are administered via computer, the storage and retrieval of information is prompt and efficient, and data are readily accessible to educators.

Computerized testing allows for the collection of additional information, such as student surveys and student motivation. Student results from an online assessment system can also be easily imported into a state-supported or district-level data warehouse. Data could be made available to schools to disaggregate their data in meaningful ways, tie to attendance rates, and other pertinent demographic variables, calculate Adequate Yearly Progress ratings, and provide teachers with up-to-the-minute student data from formative and benchmark assessments, as well as longitudinal data on student performance. Additionally, student-level information captured online prior to the assessment will likely minimize student errors in miscoding information on test answer documents.

Student Motivation

A number of researchers and field practitioners have commented upon heightened levels of student motivation when assessments are taken online. This, in turn, translates into higher performance. Also, computerized tests can measure the amount of time students spend on each test item, which has been shown to be a valid measure of student test-taking effort. In a similar fashion, student response time can be used to measure the amount of effort received by different test items. Both of these measures could be used to improve test score validity by identifying the circumstances under which students do not try very hard, and the types of test items that appear to elicit the greatest amount of effort from students. Finally, computers could be used to monitor student effort as a test is being administered, and display messages of encouragement or warning to those students exhibiting low effort. This type of effort-monitoring test has been found to yield test scores with higher validity than those from a traditional computer-based test.

COMPUTER ADAPTIVE COMPONENT TO SUPPORT-TARGETED INSTRUCTION

Moving traditional tests to computerized administration has opened the door to CAT, which is especially useful in targeting academic strengths and weaknesses of both above-grade-level and below-grade-level students. However, as explained in Component 3, a CAT system will not meet the requirements of NCLB unless all the items used for federal reporting purposes are on-grade-level and aligned to state standards. A traditional CAT system can be very useful as a formative assessment to target specific content standards/strands in classroom instruction and to informally measure students' learning.

Reduced Administrator and Instructor Effort

As stated above, if South Carolina moves to computerized testing, certain traditional test administrator activities would eventually be eliminated, including: receiving, unpacking, securing, counting, sorting, and distributing/collecting test booklets and answer sheets. However, the management functions of scheduling, monitoring, and implementing the computerized testing sessions will replace these paper-based activities.

Meeting the Needs of Special Populations

Technological advances in computerized testing signal a new era in testing for students with special learning needs and other populations, such as English Language Learners. Administering accommodated tests (e.g., read-alouds) via computer is efficient, can be tailored to the student's needs, and enhances the likelihood that standard test administration procedures are being followed.

Ability to Modify Tests

In the past, when an error was discovered within a printed test booklet, there may not have been sufficient time to correct the error, reprint testing materials, and get them where they needed to be for testing. Within the context of a computerized testing system, errors identified prior to test administration could be corrected more quickly, since the logistical aspects of printing, shipping, and distribution within the district would not be an issue.

Improved Test Security

As tests are delivered via computer, traditional handling of test materials is all but eliminated; thus, South Carolina breaches of test security may be fewer. Other types of test security issues, such as the potential of hacking into the system, need to be addressed. However, there have generally been few reports of security issues due to hacking. Sound vendor and district/school security procedures are paramount in avoiding such breaches.

One additional new issue that may lead to potential test security breaches with computerized testing is that a longer window of time will likely be needed for the test administration, since every student will not have simultaneous access to a computer to take the test. This situation creates the potential of students sharing information about the test with other students who have not yet tested. Multiple, parallel test forms delivered via a CBT system, or a modified CAT system could be used to mitigate this "downside" to computerized testing.

Barriers to Implementation

Local resistance to change— School district staff should visibly and actively support the computerized testing initiatives.

Local capacity—A school or school district may not have the technological capacity to move to computerized testing. Sufficient hardware and software lacking altogether, and the local staff may not have the technological expertise to implement and monitor a successful computerized testing program.

Mitigation of risks and obstacle to success—Due to the high-stakes nature of the tests, failure and errors should not be tolerated, and plans must be prepared in advance to deal with unforeseen situations.

FOUR KEYS TO SUCCESS

1. Communication: Frequent, brief, and effective communication is essential.
2. Involvement of technology staff: District- and school-level technology staff need to be involved from the outset of any transition to computerized testing.
3. Training workshops: Plan training sessions for all impacted school and district personnel to be conducted periodically during the school year.
4. Training materials: Online tutorials and other information, such as practice tests, for students, teachers, and test administrators, and school and district personnel are critical for the success of a CBT/CAT program.

Components 13 & 14 – A Reasonable Implementation Schedule and Conclusions and Recommendations

The recommendations presented in this section address these goals:

- the desire to receive assessment results more quickly;
- the need to address concerns about the amount of time spent on testing;
- the need to obtain diagnostic information from assessment results;
- the need to fully understand the costs associated with computerized testing; and
- the desire to provide an instructionally, psychometrically sound, and useful assessment system of the highest quality.

These goals support the need for developing a computerized assessment system that is based on solid research, will meet federal and state requirements, provides sufficient information to educators to guide them in targeting student instruction, and is fiscally prudent. Delineated below are recommendations that are grounded in the findings presented throughout this report, along with a reasonable schedule for implementation.

Recommendations for Implementation

Implement computerized testing in South Carolina using a multi-year/multi-phase rollout plan.

Since electronic media are being utilized more and more in classroom instruction, the mode of assessment must reflect the mode of instruction. However, it is critical that school districts have the capacity, such as necessary hardware, software, and infrastructure, as well as sufficient knowledgeable staff, to ensure a smooth transition to computerized testing. In order for districts to be able to reach this capacity, they must be provided with necessary funding and sufficient time to ensure adequate resources prior to full implementation. A recommended implementation plan, projected costs, and assumptions upon which the costs are based are described below. *This is a reasonably aggressive plan that may need to be implemented at a slower pace to align with available monetary and staff resources and to accommodate early*

implementation successes or challenges, evolving state and federal requirements, and state/district/school support for computerized testing.

Phase One

WHAT SUBJECT AREA?

Science

Rationale: As with any new initiative, starting off successfully goes a long way toward ensuring long-term success. Therefore, it is recommended to begin with a single subject area and that the initial subject area for statewide implementation should be one that is not currently used in NCLB Adequate Yearly Progress (AYP) calculations. Science assessment results are not included in these calculations at this time. One must remember, however, that science assessment results are used currently in South Carolina's own accountability system. Thus, the non-NCLB-subject argument may not be as strong in South Carolina as in other states.

Another reason for selecting the subject area of science for initial implementation is the fact that South Carolina has eliminated constructed-response items from its current science assessments at elementary and middle school/junior high school levels, which would greatly simplify scoring and facilitate rapid reporting. Additionally, the advice received from the Expert Panel supported starting with a lower-stakes subject area.

WHAT GRADE?

Grade 7

Rationale: As a result of the implementation of South Carolina's Act 254, pre-high school science assessments are now census tests at grades 4 and 7. Thus, initial implementation of computerized testing could most easily take place at one of these census grades. Grade 7 is recommended; since students at that level will likely have had more experience using computers in school than students in grade 4.

WHAT TEST DESIGN?

Summative CBT and Formative Assessments

Rationale: This study recommends that the initial test should be a summative CBT for Grade 7 science and should be accompanied by a computerized formative test for Grade 7 science. The Grade 7 formative science assessment could be a single formative system provided by the state or one of the Grade 7 formative science assessments from the state-approved list, with the stipulation that districts must administer one of the assessments from that list, whether that formative assessment is a CBT or a CAT. If this approach is taken, students will receive feedback about their academic strengths and weaknesses well in advance of the spring summative test so they can receive targeted instruction on South Carolina's content standards. Thus, the link between assessment and instruction will be forged.

WHOM TO TEST?

State representative sample of 1,500 students and additional volunteer districts/schools

Rationale: For purposes of state and, eventually, federal reporting, it is imperative that student test scores from summative tests be comparable within a test administration, irrespective of the mode of assessment, as well as across years, especially when the mode of administration changes. Thus, comparability studies are needed. A sample of 1,500 students per grade and subject will likely provide a sufficient number of students to conduct such comparability studies. Additionally, it is recommended that volunteer districts/schools be encouraged to participate in the computerized assessment, so that they may gain familiarity with the system and, perhaps, receive expedited test results as an incentive to participate.

Throughout each phase of the computerized testing rollout plan, a comparability study should be conducted as part of the initial administration of each computerized test.

WHICH STUDENT POPULATIONS?

General education students only

Rationale: Since special population students have varied needs with regard to assessment, it is likely that unique test forms will need to be developed, as is the current practice in South Carolina. In contrast, constructing a computerized test for the general population of

students at a particular grade calls for developing one test form. Again, with so much to be accomplished in a limited period of time, this study supports simplifying the work to be accomplished in the first phase of implementation.

Alternate Proposal for Phase 1

As a result of discussions that took place at the June 13, 2007, Advisory Committee meeting, an alternate proposal for the initial implementation phase has been included: begin with Grade 6 mathematics, rather than Grade 7 science, and then follow the same roll-out plan as described for science in Phase 2 and beyond.

Grade 6 mathematics was recommended by several committee members for the following reasons:

- Since mathematics content standards address common strands of learning across years, the results from mathematics CBTs would be very useful in informing instruction as soon as the results are received. Science standards reflect both a progression in complexity of strands across grades and topic-related strands that vary by grade, thus making results less useful for remediation purposes along a continuum of learning.
- Tools typically available for use on mathematics tests (e.g., calculators, rulers) would be an integral part of the computerized assessment, thus increasing standardization of test administration across the state.
- Since information from South Carolina's own accountability system includes test results from all four core content areas to report school and district performance, no one subject area is viewed as being more "high-stakes" than another, which is not the case with the current components of NCLB's AYP. Currently only mathematics and reading scores are used in AYP calculations.

Phase Two

WHAT SUBJECTS?

Science and Mathematics

Rationale: Based on the successful implementation of Grade 7 science in Phase 1, if that subject area and grade is chosen, it is recommended that all remaining grades of science from Grades 3-8 plus high school exit-level and end-of-course science tests that are not already being delivered by computer be administered electronically. Please note that currently the science tests at Grades 3, 5, and 7 are administered to a representative sample of students at each of those grades and not to the entire student population. Assigning specific subject-area tests to specific students is possible in some computerized testing systems through a computer algorithm. Such an automated system would greatly decrease the burden of manually assigning and distributing specific paper tests to specific students, as is the case now.

Additionally, it is recommended that one grade of mathematics should be added in Phase 2, if the plan to implement science first is followed.

WHAT GRADES?

Science at all grades and Grade 6 Mathematics

Rationale: This aggressive plan reflects a model of introducing one grade of a subject in its initial phase and adding all grades of the particular subject in the subsequent phase.

WHAT TEST DESIGN?

Summative CBT and formative assessments for all tests to be implemented in Phase 2

Rationale: This study recommends that each new assessment administered electronically should be accompanied with a corresponding formative assessment.

WHOM TO TEST?

State representative sample of 1,500 students for all new grades/subjects to determine comparability and voluntary participation of other districts/schools

Rationale: The same rationale for comparability studies needed for all added grades/subjects described above and throughout this study pertains to Phase 2, as with Phase 1.

WHICH STUDENT POPULATIONS?

- **General education students and those special education students who can take a computerized assessment using appropriate tools that are offered through the chosen test engine**

Rationale: Phase 2 could see some implementation for these students on a small scale, if the appropriateness of their participation is documented in their Individualized Education Plan. At a minimum, Phase 2 discussions should lay the groundwork for Phase 3 implementation.

Phase Three

WHAT SUBJECTS?

Science and mathematics at all grades; and English language arts at grade 6

Rationale: This recommendation supports the pattern of starting with one grade of a new subject area and fully implementing and carrying forward those subject areas begun in previous phases.

An additional consideration must be addressed, however, with the implementation of English language arts (ELA) assessments via computer. If written composition remains as part of the ELA test, policymakers must address two issues:

- Will the compositions be scored exclusively by trained human raters or will artificial intelligence be used?
- If the compositions will continue to be scored exclusively by human raters, will the ELA test as a whole be administered earlier in the school year so that scores can be reported at the same time as the science and mathematics assessments, or will the

ELA tests be redesigned as separate reading and writing assessments so that just the writing test can be administered earlier in the school year?

- If the ELA tests are split into separate assessments, will the reading portion be redesigned to include only multiple-choice items and/or short-constructed response items that can be scored using artificial intelligence or by human readers who receive the responses electronically?

These decisions must be made unless South Carolina students, parents, and educators are willing to continue to receive ELA test results following the end of the school year. In any of the scenarios for change that are bulleted above, it can be expected that the assessments will need to be redesigned, and proficiency levels will need to be reset if the test is either reconfigured or administered earlier in the school year. In addition, the redesigned tests will likely need to be submitted to the U.S. Department of Education for approval.

Paths for Implementation

Based on the scenarios for implementation of computerized testing described above, many different paths may be chosen on the way to fully implementing online testing in terms of grade and subject combinations and the order of implementation. Rather than assert that there is one single, clear direction, consider the following options:

- no change – continue on current paper-based path, or
- choose computerized testing with either science or mathematics as the first subject area to be converted to this format

CHOOSE COMPUTERIZED TESTING

Suggested Implementation Plan

Phase 1	• Science – one grade (grade 7)
Phase 2	• Science – all grades • Mathematics – one grade (grade 6)
Phase 3	Science – all grades Mathematics – all grades • English language arts – one grade (grade 6)
Phase 4	Science – all grades • Mathematics – all grades • English language arts – all grades • New subject – one grade

Rationale: This aggressive plan will dramatically accelerate innovation in the schools; utilize the most appropriate, useful, diagnostic testing in the major subject areas in order to provide results more quickly; and allow teachers to make instructional decisions based on student test results. ***This study recommends that the implementation plan outlined above should be evaluated carefully and that high degrees of success in terms of state capacity, sufficient infrastructure, and adequate staffing should be evidenced before South Carolina moves to the next phase.***

Assumptions

The following are the assumptions underlying the above implementation plan.

- A phase does not necessarily correlate to one year, and the time period for each phase does not need to be the same.
- The tests will be configured in a manner similar to the current South Carolina tests.
- A robust bank of test items exist and these items can be converted from paper/pencil format to computerized display.
- Tutorial(s) will be developed for each grade and subject.
- Electronic surveys of examinees' experiences with computers will be administered.

- Practice test(s) for each grade and subject will be developed and made available in the fall prior to each new computerized test.
- Accommodations within vendor package (e.g., highlighting, strike-through, text-to-speech, calculator, ruler, protractor) will be provided.
- A minimum testing window of three weeks, which may increase with volume of tests/students, will be needed.
- While South Carolina may choose to build its own formative assessment item bank for schools to use, *no costs for such development have been included in this plan.*
- Participation rate:
  - Per subject introduced (science, mathematics, and then English language arts):
    Phase 1 of subject = comparability sample plus 10% voluntary,
    Phase 2 of subject = comparability sample plus 50% voluntary,
    Phase 3 of subject = 90% mandatory. This formula is repeated for each subject introduced
  - Phase 1 10% participation: assumes computers at schools meet system requirements
  - Schools volunteering would test entire grade (not just selected students).
- Comparability studies would be conducted for each new subject and each newly added grade.
  - Voluntary participation (Phase 1 10%) plus representative 1,500 sample for comparability study
- This formula for implementation can become more or less aggressive by pulling forward or pushing out integration of additional subjects—high school graduation retesters; end-of-course non-NCLB, end-of-course NCLB, writing)

IMPLEMENTATION PLAN COSTS FOR BUDGETARY PURPOSES ONLY

Reasonable estimates are provided below for the implementation plan shown above. Due to the dynamic nature of the testing industry, the speed of implementation that South Carolina chooses, the possible volume of students participating, and the changing costs of technology, the

below numbers are estimates only. *Computerized testing will not result in significant cost savings.*

NOTE: Costs for bringing the state, districts, and schools to needed capacity for administering computerized tests are included in Components 7 & 9 and are NOT reflected in the implementation costs delineated below.

1. Phase 1 – $1 million
   - See Assumptions
2. Phase 2 – $4 million
   - See Assumptions
3. Phase 3 – $12/per test/per student
   - Must assume significant volume of students participating (90% at year 3 for each subject tested)
4. Phase 4 – $11/per test/per student
5. Phase 5 – $ 11/per test/per student

OTHER RECOMMENDATIONS

General

- Assessment, technology, and policy decision makers must present a coordinated effort with commitment and support from their staffs in all areas and at all levels.
- The South Carolina Technical Advisory Committee should be actively involved in the planning and implementation of computerized testing.
- The reauthorization of No Child Left Behind could have implications for large-scale assessment programs, including computerized testing, and should be tracked carefully as rollout plans for computerized testing are developed for South Carolina.
- Consider utilizing the Council of Chief State School Officers (CCSSO) Online Computer-Based Decision Making Tool.
- Evaluate lessons learned from states that were early implementers of CBT or CAT.
- Put effective back-up plans in place in case of catastrophe.

Technology

- A realistic try-out of the infrastructure should be conducted at each school to ensure that the infrastructure is adequate to handle the test itself, as well as the numbers of students who will be testing. This try-out should include consideration of time of day, day of week, and competing uses, and the number and capacity of the computer stations.
- To ensure equity of access for all students in the state, technology content standards for students should be defined and included in the curriculum.
- Specific technology content standards for both students and educators should be developed. Mastery of those standards for educators should be measured in initial licensure examinations or as a required component of educators' continuing education.
- Ongoing professional development in technology should be provided to teachers to ensure that they can lead their students in these emerging areas.

Assessment-related

- The needs of the teachers could be accommodated without compromising the accountability function with *two-stage reporting,* since preliminary data that are useful to teachers do not require final scaling or the reporting of constructed-response scores.
- Formative assessments for classroom use could include constructed-response items that are locally scored by the teacher.
- Comparability studies should be conducted during the initial year of administering each new test via computer
- Electronic delivery has the potential to provide a variety of accommodations for special needs students, and students who can benefit from the use of these accommodations should be assessed via computer as soon as feasible.

Investment

- Depending on the implementation schedule chosen, a significant investment in both infrastructure and staff will be required to ensure all schools and districts have equitable access. This investment should be sufficiently funded to lead to success.

ADDITIONAL INFORMATION

Computerized Adaptive Testing

A computerized adaptive test[1] (*CAT*) builds a customized test for each individual based on the responses to preceding items. CAT is more efficient than a fixed form assessment, using fewer items to obtain ability estimates with smaller standard errors. The item selection algorithms can be designed to satisfy all constraints of content coverage, item exposure, and security. This includes using only grade-level items and adequate sampling within content standards to provide effective diagnostic feedback.

CAT can detect and mitigate strategies for over-exposure, cheating, guessing, and manipulating the system with appropriate item selection and termination rules. Appropriate item selection and termination rules may imply alternative test delivery models, such as computer-adaptive, multi-stage testing (*ca-MST*) or *shadow tests*.

The standard recommendation concerning item bank size is approximately ten times the number of items to be administered to the typical student. However, this value depends on such factors as the distribution of ability for the students, the distribution of difficulty for the items, the diagnostic uses of the test, and the length of the testing window. One should anticipate significant start-up efforts for item development to provide adequate numbers of items across the continuum and across the content standards.

The public and the policy makers are still somewhat skeptical of individually customized test although this is not a psychometric issue. A program employing CAT will require effective communication before educators and the public will accept test results based on individually customized tests. There is little reason to consider CAT if NCLB excludes growth models and if it requires proficiency classifications be based on a common core of on-grade items. While the alternative delivery and control systems are promising, they are still being investigated. In the meantime, it is prudent to proceed slowly before implementing CAT in any high stakes situation.

[1] To date, NCLB has been reluctant to accept CAT assessments for its purposes. Specifically, it requires all NCLB testing be done with on-grade, standards-based items. It also has tended to favor fixed form assessments over individually customized tests. These issues are discussed in more detail in Component 3.

Survey Methodology

A Web survey was used to collect current information from South Carolina districts and schools about a variety of technology issues that are critical to assessing the feasibility of, and planning for, conversion to computerized testing. The Web survey provided information that was incorporated throughout several sections of the Feasibility Study Report. Separate data reports and comment reports were also produced at the state and district levels. A detailed description of all survey methods is provided in the Survey Methodology Section of the final Feasibility Study Report.

Survey topics and questions were created from several sources (e.g., existing technology surveys, internet research, subject matter experts) and refined through a multi-step, iterative process with the SC Team. A copy of the survey is included in the final Feasibility Study Report. Multiple communication channels were used to repeatedly notify districts and schools about the survey (e.g., e-mails, announcements at meetings). Both the SC and DRC Project Teams were involved in crafting communications and publicizing the importance of the survey leading up to, and during, its administration window of February 21 through March 12, 2007.

Surveys were collected for 906 schools, 20 adult education centers, 22 alternative schools, and 86 districts. A breakdown of the number of districts, schools, adult education centers, and alternative schools that responded to the Web survey is included in the Survey Methodology Section of the final report. The survey had an overall response rate of 71%.

A central goal of the Web survey was to collect data from districts and schools to provide a comprehensive picture of key issues. Due to the response rate of the survey, however, a complete set of data including every district and school was not obtained. Nevertheless, the response rate of 71% did provide enough robustness in the dataset so that it could be used to estimate survey responses for the non-responding districts and schools.

An established technique addressing with non-response in survey data was used to provide South Carolina with the statewide and district estimates that were necessary for a comprehensive feasibility study. An adjusted data record was prepared for each school that did not respond (in brief, district or state averages were used to estimate missing responses where needed). A more detailed description of the procedure is given in the Survey Methodology Section of the final Feasibility Study Report. The adjusted data set provided results for the

report sections that require a comprehensive picture at the state and district levels (Components 6–9).

- Data reports were produced for the state and districts (based on the original, not adjusted data set). A series of comment reports were also produced for the state and districts. One set of these reports includes an analysis of the major themes within the comments provided. The primary themes that emerged (for schools) on perceived barriers and advantages were as follows:
- **Biggest perceived barriers to implementing computerized testing – Technology** (e.g., technical readiness of schools, need to upgrade equipment, having enough computers and sufficient bandwidth); Space (e.g., managing/having enough space for testing, number/size of computer labs); Personnel (e.g., having enough and the right type of personnel, availability of technical staff, teacher familiarity with technology, test monitors).
- **Biggest perceived advantages to implementing computerized testing – Feedback** (e.g., receiving scores sooner, being able to use test feedback for planning); Improvements in testing process (e.g., streamlining the testing process, easier materials collection, lower amount of paperwork); Effect on students (e.g., students having an interest in and positive perceptions of computer use, students expected to perform better).
- A complete summary of the theme results is included as an appendix in the full report.