Computerizing Statewide Educational Assessments in Minnesota:
A Report on the Cost and Feasibility of Converting the *Minnesota Comprehensive Assessments* to a Computerized Adaptive Format

The Office of Educational Accountability, University of Minnesota
June 15, 2005

What is Computerized Adaptive Testing?

- Test in which items are administered by computer and adapted to the achievement level of the individual test-taker
- High achieving students receive successively harder items, including above grade level items, to establish the level of material mastered
- Low achieving students receive successively less difficult items, including below grade level items, to establish the level of material mastered

Introduction

- Statewide testing will eventually be computerized so the key question is whether it should be adaptive.
- We assume Minnesota wants to implement a system that is compliant with No Child Left Behind.
- If Minnesota were to decide to update the *MCA*s with a computerized adaptive test, an RFP would be issued and open bidding would occur.

Methodology

- Personal interviews with teachers, testing directors, technology staff, principals, legislators, and test vendors
- Focus groups with teachers, parents, education associations, and school personnel
- Online surveys of school technology staff

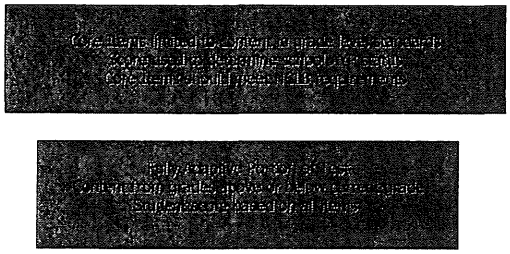
Possible Advantages of Computerized Adaptive Testing

- More precise estimates of achievement (lower standard errors)
- More accurate student growth measures in a value added system
- Quicker turnaround time for results
- More diagnostic and instructionally informative
- National comparisons

Possible Disadvantages of Computerized Adaptive Testing

- Longer testing window for schools
- Possibly more items per test
- Loss of instructional time on computers during testing window

Meeting NCLB Requirements with an Adaptive Test



Feasibility

- Online survey of school technology capacity
- Responses were analyzed to determine if schools had the capability to do computerized testing based on minimum technological requirements

Feasibility

- Almost 80% of the schools in our survey seem able to administer a computerized adaptive test with their current configuration or with a modest alteration in the configuration if
  - The testing window is lengthened to 4 – 5 weeks
  - Constructed response items are administered paper/pencil or dropped initially until students can be expected to compose on a computer
  - During a transition period (say 3 years), some schools that lack computer capacity can administer paper/pencil

Cost

- Annual Operational Costs of Test Administration for Reading and Mathematics Grades 3 – 8, 10, 11
  - $8.5 – 9 million without constructed response
  - Substantially increased with constructed response: $21 million (estimate)
- School technology: $30,000 to $35,000 per lab
  - Student workstations
  - Servers
  - Networking

Recommendations

- In order to ensure comparability, attention must be paid to the presentation of reading passages and reading items on the screen, opportunity to review items, and opportunity to highlight material.
- We recommend lengthening the testing window to 4-5 weeks to complete all exams in all grades. We recommend that MDE require schools to test students in a given subject and grade with a two week testing window.

Recommendations

- We recommend that the computer adaptive tests be administered in reading and mathematics only in grades 3 -8, reading grade 10, mathematics grade 11.
- To set the stage for computerized writing assessments, we recommend content standards be established stating a grade by which students are expected to be proficient in composing essays and constructed response answers.
- We propose that the test begin with on grade level items aligned with the state standards administered in a fixed or adaptive format followed by a fully adaptive portion.

Recommendations

- If Minnesota adopts a computerized adaptive testing format, for reasons of cost, testing time, and logistical feasibility, the state should consider using only multiple choice items initially. Constructed response items could be added when students can enter their responses on the computer.
- Revise the information required of vendors through the request for proposal process: i.e. adding criteria on hardware/software requirements, diagnostic materials, experience with large-scale computerized adaptive assessments, and training provided for interpreting results.

Recommendations

- In developing reading test, Minnesota should consider:
  - Shortening reading passages such that the entire passage can fit on a single screen.
  - Employing a simulated "page turning" technology in which student do not scroll through a passage, but rather click a button to go to the next page.
  - Allowing highlighting of key phrases

Recommendations

- Computerized testing is coming to statewide testing in Minnesota. It is already here in many districts. It remains to be decided when and how Minnesota will make the transition to computerized statewide testing and whether the computerized assessments will be adaptive.

Proposed Implementation Schedule

- 2005 – 2006: Draw up and issue request for proposal and choose vendor.
- 2006 – 2007: Draft items in spring and fall 2006. Field test in 2007. Create sample tests that schools and students can access in spring 2007.
- 2007 – 2008: Begin administering computerized adaptive tests in spring 2008.

Computerizing Statewide Educational Assessments in Minnesota: A Report on the Cost and Feasibility of Converting the *Minnesota Comprehensive Assessments* to a Computerized Adaptive Format

June 2005

Office of Educational Accountability
College of Education and Human Development

UNIVERSITY OF MINNESOTA

Computerizing Statewide Educational Assessments in Minnesota:

A Report on the Cost and Feasibility of Converting the

Minnesota Comprehensive Assessments

to a Computerized Adaptive Format

has been produced by the:

Office of Educational Accountability (OEA)

College of Education and Human Development

University of Minnesota

Authors

Kristin A. Peterson, *Office of Educational Accountability*
Mark L. Davison, *Office of Educational Accountability*
Leah Hjelseth, *Office of Educational Accountability*

Advisory Committee

Jim Angermeyr, *Bloomington Public Schools*
Tim Hodges, *Farmington Public Schools*
Robert Kochmann, *Sauk Rapids Public Schools*
Dirk Mattson, *Minnesota Department of Education*
David J. Weiss, *University of Minnesota*

This report was written in collaboration with the Minnesota Department of Education. Special thanks are due to the members of our advisory committee for their time and efforts. The views expressed in this report are those of the Office of Educational Accountability.

Computerizing Statewide Educational Assessments in Minnesota: A Report on the Cost and Feasibility of Converting the *Minnesota Comprehensive Assessments* to a Computerized Adaptive Format

June 2005

Office of Educational Accountability
College of Education and Human Development

UNIVERSITY OF MINNESOTA

For these and other reasons, it seems inevitable that paper/pencil tests will be replaced by computerized tests at some point in the future, and indeed the Minnesota Department of Education (MDE) has announced plans to implement computerized testing beginning in 2009. However, the computerized adaptive tests currently administered by Minnesota schools would not satisfy the federal testing requirements of the No Child Left Behind Act in their current form and therefore would jeopardize federal funding for education in Minnesota. Furthermore, computerizing tests can involve adjustments to the computer medium. Those adjustments can involve changes in test content, format, and/or administration procedures.

Before committing to a computerized adaptive administration of the *Minnesota Comprehensive Assessment* (*MCA*), it was deemed prudent to conduct a cost and feasibility study of the changes that would be required and the consequences that may result. The Office of Educational Accountability, in cooperation with the Minnesota Department of Education (MDE), was charged by the Minnesota legislature with the responsibility of conducting the feasibility study (Omnibus K-12 Education Policy Bill, 2003-2004). This report contains our findings concerning the issues, cost and feasibility of converting the *MCAs* to a computerized adaptive format.

This report begins by stating the major purposes of statewide testing in Minnesota and stipulations that the testing program must meet. Minnesota's assessment and accountability systems are governed by federal and state statute both of which have heavily influenced our recommendations. The report then proceeds to a description of the procedures used to gather information for this report. It then provides a review of the literature on the comparability of scores obtained by examinees when assessments are administered on computer rather than paper/pencil. Next, the report provides background information on Minnesota's current assessment system and upcoming changes in that system. Then we describe our recommendation for a computerized adaptive testing system that strives to satisfy the various considerations described in earlier sections. Finally, we discuss the feasibility and costs of such a system.

While our legislative charge did not call for the development of a detailed plan, it is difficult to develop a cost estimate without having a plan on which to base those cost estimates. Therefore, our recommendations include a computerized adaptive test

administration plan on which the cost estimates are based. The plan does include options that will affect the cost, but the plan is a theme with variations, not several distinct testing plans.

Purposes of Testing and Governing Legislation

The Minnesota Department of Education states the following purposes for the new *MCA-II's*, purposes with which we readily concur:

1. The purpose of the *MCA-II* is to document, or measure, Minnesota students' achievement with regard to Minnesota Academic Standards.

2. The *MCA-II* results can be used to inform curriculum decisions at the district and school level and inform teaching at the classroom level through the use of classroom level reporting.

3. The results can be used to help demonstrate growth from year to year for students or groups of students using both cohort and longitudinal student data. (MESPA Institute 2005)

As we prepared this report, we considered a number of desirable characteristics for any statewide educational testing system:

- The content should be aligned to state academic standards;
- The assessments should satisfy federal NCLB requirements so as not to jeopardize federal funding;
- Cost to the state and to local school districts should be reasonable;
- Delivery of results for schools and students needs to be timely;
- Results should be instructionally informative;
- The hardware, software, and staffing requirements should be feasible for schools across the state;
- It must be fair for all students and schools in the system;
- It should be compatible with school and district technology; and
- It should keep total testing to a minimum.

Given these purposes for testing, making Minnesota's computerized tests adaptive will only make sense if the adaptive tests measure students' achievement with regard to Minnesota Academic Standards, improve measures of student growth, and enhance the informative value of results for district, school, and classroom use.

Issues to Address with a Computerized Testing System

Since adoption of statewide testing in Minnesota in 1996, there have been at least six recurring issues that may be addressed through computerized adaptive testing itself or by embedding the testing within a larger norm-referenced and/or diagnostic testing system. We want to be clear on one major point. Our recommendations go beyond simply recommending that the *MCA-II* be administered in a computerized format. We are recommending that the *MCA-II* be administered in a computerized adaptive format *and* be embedded within a larger norm-referenced and/or diagnostic assessment system. Of course, districts would still have the option of continuing their current norm-referenced testing and diagnostic testing. However, if this recommendation is adopted, the Minnesota statewide tests used for accountability purposes may include national norms and/or diagnostic information. Furthermore, if this larger change is adopted, then the new testing system would, to varying degrees, address six recurring issues: minimum competency vs. more challenging assessments, national comparisons, growth assessment and value added school evaluation, classroom utility of assessment information, and minimizing testing in schools.

Finally, Minnesota has wanted to keep the total testing time in schools to a minimum. Currently, schools administer additional tests in order to obtain diagnostic information and to obtain national norm information. If the statewide testing were to provide diagnostic information or national norm information, then some districts may decide they no longer need to administer an additional test in the spring in order to obtain diagnostic or national norm information. The number of test administrations in the districts would thereby be reduced.

Adaptive testing involves giving each student a unique set of items tailored to his or her achievement level. Therefore, we propose that the adaptive test begin with on-grade-level items aligned with the state standards administered in a fixed or adaptive

format as permitted by the United States Department of Education (USDE). The on-grade-level portion would be followed by a fully adaptive portion. Each student would take a unique set of items, but within that unique set of items, there would be a subset covering content from the Minnesota standards at the student's current grade level. For the purposes of the federal accountability system, schools would be evaluated on how well their students performed on grade level items only. Hereafter, we refer to the on grade level items as the Core Items.

Federal requirements for statewide assessment programs are outlined in the 2002 reauthorization of the Elementary and Secondary Education Act, the No Child Left Behind (NCLB) legislation (No Child Left Behind, 2001). Assessment guidelines issued by the U.S. Department of Education in conjunction with NCLB require tests aligned with standards[1] and forbid use of out-of level tests in the federally mandated evaluation of schools (Standards and Assessments, 2004, p. 2). An adaptive test would be an out-of-level test if some of the items administered were above or below the student's current grade level.

The intent of this opposition and its effect on adaptive testing is unclear. The U.S. Department of Education (USDE) may intend that all students at a given grade in a given year take exactly the same items. If that is the case, the Minnesota computerized adaptive test would have to include a common set of items given to all students before beginning the fully adaptive portion of the test. Alternatively, the USDE may have only intended to preclude items from lower or higher grades. In this latter case, Minnesota may be able to satisfy NCLB requirements by starting with a limited adaptive test in which all items cover content from the student's current grade followed by a fully adaptive portion in which content may come from lower or higher grades if appropriate. This latter option is preferable, because it would shorten testing time. In either case, Minnesota should be able to satisfy NCLB requirements by beginning the test with an on grade level set of items, the same items for all students or adaptively administered,

[1] USDE requires that the items cover content in the state standards *for the given grade of a student.* That is, a third grade student must take items covering content contained in the state's third grade standards. An item on the third grade test that covered content in Minnesota's fourth grade standards or Minnesota's second grade standards would not be considered an aligned item, because items on the third grade test must come from third grade standards.

followed by a fully adaptive portion that would include items from grades above and below that of the student if appropriate.

If Minnesota were to replace the current *MCAs* with a computerized adaptive test, it would presumably purchase that test through one of the national vendors. Minnesota statute/rule requires that such a large purchase occur through an open bidding process. A request for proposals would be issued and open bidding would occur. This means that although a number of school districts around the state are currently administering computerized adaptive tests, adopting a statewide system would not necessarily mean expanding what these districts are doing.

The legislative language refers to "replacing" the current testing system. Rather than replacing the *MCAs,* what we propose is computerization and augmentation of the current testing system. The *MCAs* have been developed as an accountability tool to satisfy federal requirements under NCLB. Barring a change in federal requirements for state testing programs, the best way to develop a computerized adaptive testing program satisfying federal requirements is to incorporate a core of items aligned with grade level expectations into the larger computerized adaptive test.

Methodology

In compiling this report, we relied on several sources of information, including personal interviews, meetings, focus groups, and an online survey.

Interviews and Focus Groups

We began by interviewing legislators and legislative assistants who were involved in the initiation and development of this proposal. We spoke with them about the background of the bill and the intentions for moving forward. We also informed them of how we intended to carry out the assignment and with whom we planned to speak.

We continued our interviews with key personnel from several school districts around the state that were currently administering some form of computerized testing in some capacity in their schools. These interviews included principals, superintendents, testing coordinators, and technology staff. Participants were asked about advantages and disadvantages of their current computerized assessment systems regarding test

administration, turnaround time for results, reporting capabilities, and costs. Other questions centered on the technology infrastructure in their schools and districts and student workstations being used for testing. They were also asked which grades of students in their schools took the tests via computer and their opinion of a strictly multiple-choice format. Finally, test security concerns and protocol were discussed.

Second, we met with five selected testing vendors about their computerized testing products available now or in the near future. They were asked to describe their computerized testing products and to indicate whether those products included the capacity to adapt the tests. They were asked questions about test development and delivery, staff training, and reporting. Vendors also provided information on the minimum hardware and software requirements of their systems. Approximate estimates of cost were discussed, although precise estimates would of course depend on the exact nature and number of tests to be administered. Vendor estimates form the basis of cost estimates later in this report (pp. 35–37). The minimum hardware and software requirements of two vendors were incorporated into survey questions described below in order to estimate the percentage of Minnesota schools that could meet those technological requirements. Appendix B contains a set of questions that was emailed to vendors before the interview.

Third, we conducted numerous focus groups with parents, teachers, and education associations and groups across the state. Participants were asked about their experiences with the current state testing system, as well as their thoughts on the proposal of switching to a computerized adaptive testing system. They were asked about other school-wide or district-wide tests that they were administering and the instructional informativeness and cost of those assessments. They were also asked about the possibility of dropping constructed response items if that became necessary to facilitate conversion to computerized testing, as well as the feasibility of lengthening the testing window, security issues that might arise with computerized testing, and students' abilities to compose answers to constructed response questions using a keyboard. Appendix C contains the basic set of questions posed in these interviews, although the discussion took somewhat different directions with each group. Many of the questions were the same as those asked of school district personnel with some adaptation to the composition of a

particular group. For instance, we did not ask parents about the technology capacity of their child's school.

Finally, we had meetings with MDE staff and staff at the legislative auditor's office for clarification of the current assessment system and budgeting issues. For a complete list of the people and groups who were interviewed for this study, please see Appendix D.

Technology Survey

In addition to interviews and focus groups, a survey was conducted to assess the current technology in schools. The survey was completed by technology staff and reflected their perspective, rather than that of teachers or administrators. The current 2004 minimum hardware and software requirements provided by vendors served as the basis for several of these questions.

The technology survey was developed by the Office of Educational Accountability in collaboration with an advisory committee. The advisory committee was made up of Office of Educational Accountability (OEA) staff, a University of Minnesota faculty member specializing in computerized adaptive testing, a Minnesota Department of Education (MDE) staff member, a school media specialist, a teacher and former technology coordinator, and a district testing director. While the advisory panel gave us valuable input, the content of the survey was the responsibility of OEA.

The purpose of the survey was to gather data on the current technology capacity in Minnesota schools. The survey consisted of 29 questions and was administered through an online survey facility. An email message, along with a URL link to the survey, was sent to superintendents and technology directors in districts and schools around the state by the Minnesota Department of Education. The topics of the questions included: the number and types of computers that schools have for student use; computer networking; internet connections; school and district server capacity; and technology staff. Data were collected and analyzed by the Office of Educational Accountability. (See Appendix E for a copy of the survey.)

Data were collected from over 500 schools around the state via the online survey. The St. Paul school district provided its data via an Excel file that was later merged with

the data from the online responses. The Minneapolis school district provided one survey for elementary schools, one for junior high/middle schools, and one for high schools with the average numbers and calculations. Therefore, for Minneapolis schools the same data were entered for every elementary school in the district. The same is true for junior high/middle schools and high schools. In total, data were collected representing approximately 700 Minnesota schools, approximately 50% of schools statewide.

Several hardware and software questions were asked based on the minimum hardware and software requirements of two vendors. Because these questions pertained to very detailed components of technology, they may not have always been understood as intended by respondents. Some schools may have the technology capacity but may operate under different programs or specifications that were not specifically mentioned in the survey. Other schools are operating with more advanced systems than mentioned in the survey, and may not have understood the specifications in the survey as minimum requirements. While these limitations could lead to over- or under-estimation of the number of schools that can administer these tests, we think they are more likely to lead to an underestimation.

At least four other limitations of our survey methodology should be noted. First, if computerized testing is implemented and phased in over several years, it is possible that the minimum specifications for computerized testing may rise, in which case some of the current systems would be too old or out-of-date. Second, if some of the systems are not replaced by the time computer-based testing is implemented; they may meet the minimum specifications but because of age be more susceptible to breakdowns during testing than at the time of the survey. Third, we asked only about the capacity needed to implement simple computerized adaptive testing that includes items similar to those now administered on a paper/pencil test. We did not ask about the capacity to administer cutting edge items that involve sophisticated graphics. From our results, readers cannot draw conclusions as to the number of schools that could administer computerized items involving more advanced graphics, video streaming, or media files.

Literature Review:

Comparability of Computerized and Paper/Pencil Assessment Results

In Minnesota's statewide testing, the comparability of computer-administered and paper/pencil-administered tests is important for two reasons. First, if in the early stages of computerized testing, schools with inadequate computer facilities are unable to administer the tests via computer, some students may be taking a computer-administered form of the test while others are taking a paper/pencil version. Even if most students in every school take a computerized version of the test, some students with disabilities may still need a paper/pencil version (e.g., blind students needing Braille). Therefore, at least in the early stages, Minnesota may have a dual system with some students taking a computerized form of the test and others taking a paper/pencil form. In such a dual system, it is important that neither the computerized nor the paper/pencil version provide an advantage to one group of students over the other.

The second reason why comparability is important is that Minnesota tracks the performance of schools over time. For example, in determining whether a school has made safe harbor under the Adequate Yearly Progress (AYP) evaluation of No Child Left Behind, the school's proficiency index from the current year is compared to the proficiency index from last year. When Minnesota makes the transition to computerized testing, the school's performance on the current year's computerized test will be compared to its performance in prior years under paper/pencil administration. For these longitudinal comparisons to be valid, results from paper/pencil and computerized administrations must be comparable.

Because comparability of results is important, we examined the research literature for studies comparing results from the two modes of administration.

General Comparability

Several studies have compared results from computer- and paper/pencil-administered tests. Because Minnesota's tests are untimed, this short review will only cover studies of untimed test administrations. In general, the literature supports the near equivalency of untimed computer based tests and paper/pencil tests *when the computer provides opportunities comparable to paper/pencil testing conditions*. Mazzeo and

Harvey (1988), Mead and Drasgow (1993), and Russell, Goldberg and O'Connor (2003) reviewed the existent literature. Mazzeo and Harvey concluded that computerized forms yielded lower scores on untimed tests, but the differences were minimal. Five years later, Mead and Drasgow's review found an effect for media, with computerized tests being slightly harder; but again the effect was not significant. It should be noted that much of this literature is based on young adults rather than children in K–12 schools.

A recent study by Poggio, Glasnapp, Yang, and Poggio (2004) exemplifies the studies in the three review articles cited above, a study applicable to comparability of results in statewide testing. Their results and conclusions seem rather typical. Poggio, et al. compared four groups of students taking the mathematics portion of Kansas' statewide assessment. In their study, students took the mathematics test twice, once administered by paper/pencil and once administered by computer. Some students took the paper/pencil version first while others took the computer version first. While students taking the computerized version did score slightly lower on average, the differences were small and not statistically significant. Based on the analyses reported, results "make very clear that there existed no meaningful statistical differences in the composite test scores attained by the same students on a computerized fixed form assessment and an equated form of that assessment when taken in a traditional paper and pencil format." (Poggio, et al., p.14)

If the testing conditions are not comparable under both forms of administration, however, then results for the two types of testing may differ more substantially. For example, if students taking the paper/pencil version have greater opportunity to review and correct prior answers, then students taking the paper/pencil version may perform better. In reviewing the literature on comparability, we will highlight testing conditions that need to be comparable if results are to be comparable.

Given the nature of Minnesota's current paper/pencil tests, at least three factors need to be considered in converting to computerized testing: scrolling of reading passages, reading passage length, and review of items/answers.

Scrolling, Item Presentation, and Passage Length in Reading

In reviewing comparability of results on reading tests, we examined some studies in the testing literature but also some in the reading literature. These reading studies

compared people's comprehension when text was presented on a computer screen as opposed to paper.

Kiely, Zara, and Weiss (1986), testing members of the armed services, found that reading scores on the paper/pencil test were higher than those on three computerized versions of their reading test. Most important, for our purposes, mean scores on the computerized test were more comparable to those on the paper/pencil test when examinees were able to scroll to different sections of the reading passage with the item remaining in view.

Dyson and Haselgrove (2001) reviewed several studies suggesting that examinees—primarily between the ages of 18 and 24—read more slowly from computer screens, something which may not be important on an untimed test. More important, for our purposes, is their finding that line length influenced comprehension on the screen with medium line lengths, roughly 55 characters per line, being optimal.

Several studies found no difference in reading comprehension between the computer screen and printed text (e.g., Gambrell, Bradley, & McGlaughlin, 1987; Rice, 1994, Reinking, 1998), although reading from the screen seemed to be slower. In those that found a difference in reading comprehension, the difference did not consistently favor the computer or paper/pencil versions (e.g., Kiely, Zara, & Weiss, 1986; vs. Meyer & Poon, 1997). In studies that did find better comprehension from paper/pencil, at least on some types of items, the authors conclude that scrolling may be the source of the difficulty (Kiely, Zara, & Weiss, 1986; Dyson & Haselgrove, 2001).

Highlighting Content in Reading Passages

In Minnesota's current *MCA* reading tests, students are allowed to highlight or underline text in the reading passages so they can readily refer back to important text (e.g., main idea) as they answer questions. If passages are short, highlighting is unimportant, because students can readily find key phrases. If passages are long, then highlighting may aid comprehension (Rice, 1994).

Item Review and Changing Answers

In a paper/pencil test, students can return to a completed item, review a previous answer, and change that answer if desired. As a rule, computerized adaptive tests do not allow students to return to an item and change an answer. Computerized non-adaptive tests can allow students to review and change answers, although sometimes students are allowed to review answers after all items have been completed, but not before.

Does denying students the opportunity to review answers affect their scores? Studies by Lunz (Lunz, Bergstrom, & Wright, 1992; Stone & Lunz, 1994) concluded that students increase their mean number of items correct when they are allowed to review and change answers. The overall effect is not large, however, because students do not change many answers, and sometimes the change is from a correct item response to an incorrect item response.

Overall, the research suggests that computerized testing results and paper/pencil results on the same students correlate highly. Mean scores on the tests seem to be nearly equivalent, but there is a trend toward minimally lower means on computerized forms. We cannot rule out the possibility that these small mean differences are due to some incomparability in the administration of the tests. The differences are sufficiently small that some researchers (e.g., Poggio et al., 2004; Mead & Drasgow, 1993) have concluded that the differences can be ignored. Indeed, at least two statewide testing programs have ignored the differences. Both Idaho and Oregon now offer or have offered schools the opportunity to administer their statewide exam by computer or paper/pencil and have treated results from the two administration forms as equivalent. To make results as comparable as possible, test construction must attend to the presentation of text and items in reading assessments, student opportunity for review of item responses, and student opportunity to highlight key text in long reading passages.

Minnesota's Current Statewide Assessment System

Currently, the statewide assessment system in Minnesota consists of several components. The first component is the set of high school graduation tests, the *Minnesota Basic Skills Tests (BSTs)*. The second component is the set of tests aligned with

Minnesota Academic Standards in reading, writing, and mathematics, the *Minnesota Comprehensive Assessments (MCAs)*. In 2008, science tests will be added to the *MCAs*. Finally, Minnesota has a set of tests to measure the English language development of second language learners, the *Tests of Emerging Academic English (TEAE)*.

The charge to our office covered only the *MCAs*. However, our recommendations below need to be considered in the context of other tests. Specifically, MDE is in the process of computerizing the *TEAE* and *MCA* science tests. Our estimates of computer lab time lost to instruction through testing include only time lost for administration of the reading and math tests. Additional computer lab time will be taken from instruction by the *TEAE* and *MCA-II* science assessments as they are also computerized.

We have not considered the science tests in our recommendations below. The Minnesota Department of Educations (MDE) has presented an exciting proposal for new computerized science tests that incorporates advanced graphics into large scale assessment. This test will employ new types of items that are impossible with paper/pencil testing. We look forward to the day when such technology advances are incorporated into assessments in every subject area. However, since the technology requirements are much greater for tests of this nature than for the proposed reading and mathematics assessments we have not included the science tests in the proposal.

We have also not included the writing tests. Focus group discussions with school staff indicated that each school is free to determine the grade by which students are expected to compose written material using a keyboard, rather than paper/pencil. Thus, there is currently no grade at which one can say that it is reasonable to expect all students across the state to take their writing tests on computer.

The focus in this report is on reading and mathematics assessments. These assessments are aligned with state academic standards. Tests are administered in grades 3, 5, 7, 10 (reading only), and 11 (math only). Beginning in 2005–06, there will be tests added in grades 4, 6 and 8, so that testing will occur in both reading and mathematics in grades 3–8, reading only in grade 10, and mathematics only in grade 11. The tests are administered via paper/pencil during a designated three week testing window. Schools are allowed to decide when they will test their students during that three week time frame.

The Minnesota Department of Education and the testing vendor begin writing a test plan and items for an assessment approximately twenty-four months before an assessment is administered. Initial draft items are reviewed for content and bias in the summer prior to field testing. The items are field tested one year before they are administered as operational items. For instance, items intended for use in Spring of 2005 would be drafted and selected in the spring of 2003, reviewed for content and bias in the summer of 2003, printed as field test items in the fall of 2003, administered as field test items in the spring of 2004, and incorporated as operational items in the spring test of 2005.

Once final items are selected for the operational test, the test booklets are printed and delivered to schools. School personnel administer the test, keeping track of booklets and answer sheets through the administration process. Before and after students complete the test, the answer sheets and test booklets are kept in secure locations at the school until all students have taken the test. Once all the students in the school have completed the test, the tests and test booklets are picked up at the school and delivered to the testing vendor where the tests are scored and stored. School staff in our interviews and focus groups argued that the current system cost them large amounts of staff time and salary to organize, administer, and monitor the tests in their current form.

As stated above, once reading and mathematics testing is in place for grades 3–8, 10, and 11, the reading and mathematics testing alone will involve printing, delivering, and returning approximately 880,000 test booklets and 880,000 answer sheets each year. If tests are fully computerized, these printing and delivery processes will be replaced by electronic delivery that saves staff time at the district level. At the state level, computerization can save costs associated with printing, delivering, returning, scanning, and scoring test booklets and answer sheets. A few accommodated forms (e.g., Braille forms) cannot, to our knowledge, be delivered via computer and will continue to be administered via paper/pencil. Some of these savings in time and costs may be partially offset by the need for increased technology equipment, maintenance, and support, but some savings should be realized.

Once test booklets and answer sheets have been returned to the vendor and the answers are scored, results are sent to the Minnesota Department of Education as well as

schools and districts. At present, there are separate score reports for students (and their parents), the school, and the district. Starting with the *MCA-II*, there will also be a score report for each classroom. Schools receive results approximately 8–10 weeks after completing the tests.

Whether the current system minimizes testing is debatable. Some districts administer supplemental diagnostic or norm-referenced tests to meet the needs of their students. Other districts rely heavily on the *MCA*s and forego additional testing—primarily for reasons of cost and testing time.

In reading our recommendations below, particularly our schedule for implementation of computerized adaptive testing, readers need to remember the sequence of events described above that needs to take place before an item becomes part of an operational test.

Recommendation of a Computerized Adaptive Testing Plan: A Theme with Variations

In order to examine the feasibility and cost of computerized adaptive testing, a plan on which the costs and feasibility estimates are based must be developed. In this section, we describe the plan with some variations. In the following two sections, we discuss the feasibility of and the costs associated with the plan. Following our legislative charge, the plan covers only the *MCAs* and does not cover *BSTs* or the *TEAE* tests. There are eight key elements of the plan itself, around which this section is organized:

- Subjects and grades to be tested
- The testing window (e.g., the schedule for testing)
- Computerizing and adapting so as to satisfy NCLB requirements
- Embedding the statewide tests in a larger assessment system
- Constructed response items
- Length of reading passages
- Selecting a vendor
- Implementation schedule

Subjects and grades to be tested

We are recommending that computerized adaptive testing begin with reading and mathematics in grades 3–8, reading in grade 10, and mathematics in grade 11.

An objection can be raised to our recommendation to implement a computer adaptive format for reading and mathematics assessments since these are high stakes tests for schools under NCLB. A more cautious approach would be to begin computerization in science and/or writing because these tests are not high stakes for schools and because they are not given in every grade. However, we have chosen these two subjects for several reasons.

First, reading and mathematics lend themselves to computerized adaptive testing as evidenced by the fact that much of the research comparing computerized and paper/pencil testing is based on reading and mathematics assessments. If growth indicators are implemented in a value-added system, those indicators will most likely be indicators in reading and mathematics. One advantage of computerized adaptive testing is that it provides more accurate assessment, particularly at the high and low achievement extremes, and thereby provides a more accurate measure of student growth. Because computerized adaptive testing provides better measures of student growth and because a value-added system based on student growth will most probably be implemented in reading and mathematics, we are recommending implementation of computerized adaptive testing in those two subject areas.

Furthermore, because reading and mathematics are high stakes content areas for schools, it is imperative to make good, instructionally informative data available to teachers in those subject matters. By providing reading and mathematics test results to teachers and administrators in a more timely fashion, faculty can more readily use statewide test results in planning reading and math instruction for the following year in the high stakes subject areas in which the information is most needed.

Writing tests, at least those that demand samples of student writing, do not lend themselves as easily to a computer adaptive format. Furthermore, administrators and teachers told us that there is no consistent grade at which students are expected to be able to compose prose using a keyboard. While we are recommending against computerizing writing exams initially, we look forward to the day when students will take writing exams

on computer. Therefore, we recommend that Minnesota amend its academic standards in writing to establish a grade level by which students are expected to compose text passages on computer. Once the amendments to the standards have been implemented, then writing assessments can be administered via computer.

In science, the Minnesota Department of Education (MDE) is planning computerized tests that employ interactive items with advanced graphics and video streaming. To our knowledge, no large-scale computerized adaptive testing program to date has incorporated such items. Therefore, we think it prudent to delay a decision about computerizing the science test items in an adaptive fashion.

Testing Window

With paper/pencil administration, several classrooms of students can take the test at the same time. Many schools, particularly elementary schools and small secondary schools, have only one computer lab. This means they could administer the tests to only one classroom at a time. Therefore, if computerized adaptive testing is adopted, the time interval in which tests are given needs to be expanded.

Currently the testing window for administering the *MCA*s is three weeks. That is, schools have three weeks in which to administer the tests. Given the technology available in schools, the testing window would need to be lengthened to allow approximately four to five weeks for students to complete all exams in all grades. In most schools, one exam (i.e., one subject in one grade) could be completed in a single day. To prevent students in some schools/districts from having more time to learn material, MDE should require all schools to administer the tests for a given grade and subject within a two week testing window. That is, the tests for any one grade and subject would be given within a two week window and tests in all grades and subjects would be given in a four to five week testing window.

Administering the test to one classroom at a time can potentially create security problems in schools that have more than one class at each grade. For example, after the first classroom of third graders has completed the test, but before the other third graders have taken it, the third graders who have completed the test may talk with students who have not yet taken the examination. Elementary and junior high teachers with whom we

talked did not seem very concerned about this potential problem. However, there was a fair amount concern among high school teachers and administrators about the possibility of students discussing the test with one another.

The longer testing window also means that for up to four or five weeks, the school's computer lab(s) may be unavailable for instructional use. This is equivalent to approximately 11% of the school year. Additional computerized tests, such as the computerized *MCA* science test and the *Test of Emerging Academic English* would add to the computer lab time devoted to testing.

In focus groups, administrators and teachers experienced with computerized testing have reported that teachers often initially objected to giving up the instructional lab to testing. However, after one or two years experience with computerized testing, the objections dissipated if teachers saw that the computerized testing produced more instructionally informative results. If computerized testing will result in substantial loss in computer lab time for instructional purposes, it is imperative that results be useful to teachers. To avoid this loss of instructional time, Minnesota should increase the number of computer labs available in Minnesota schools.

Satisfying NCLB requirements: Core and Adaptive items

As mentioned earlier, initial informational materials by the U.S. Department of Education (Title I Directors Conference, 2003 www.ed.gov/admins/lead/account/standassess03/edlite-slide17.html, www.ed.gov/admins/lead/account/standassess03/edlite-slide23.html) required a uniform testing system for all students and opposed out-of-level testing and thereby seemed to oppose the use of a computerized adaptive test for purposes of evaluating underperforming schools (determining AYP). Their exact intent is unclear. The law does clearly specify that tests used for determination of AYP must be aligned with state grade level standards; i.e., must cover content included in the standards for the students' assigned grade. Guidance given at the 2003 Title I Directors Conference says that states must use a uniform set of assessments statewide or a combination of state and local assessments. Further, that guidance goes on to say that states cannot use out-of-level tests, meaning tests that cover standards above or below the student's current grade level.

While the exact intent of this guidance is unclear, the intent may have been (a) to require that all students in a given grade in a given year take exactly the same items, or (b) that all items taken by a student must be drawn from the content in the state standards at the student's current grade level. Either way, Minnesota seemingly could satisfy the NCLB requirements by administering a computerized adaptive test, a portion of which included only items covering the standards in the students' current grade.

If the intent is to require all students in a given grade in a given year to take exactly the same items, then Minnesota would have to first administer a common set of items aligned with grade level standards (the grade level core) followed by items administered in an adaptive fashion. A school's AYP status would be determined solely on the basis of the grade level core. The adaptive items at the end would give a more precise estimate of student achievement levels, particularly for students at the high and low ends of the achievement continuum. Student scores would be based on both the adaptive and grade level core. Combining a common grade level core and adaptive items, however, would yield a test with more items than the current *MCA*s.

Since not all students take the same number of items in an adaptive test, it is difficult to say exactly how many more items would be administered. Based on the experience of another state (Idaho), if students have to take a common grade level core in addition to the adaptive portion we would predict that some students would take about 50% more items than with the current *MCA*s, while other students may take no additional items. Testing time would not increase by 50% on average, however, because the additional items would all be multiple choice. Whether the additional items would require additional time depends on whether the test includes constructed response items as currently required by Minnesota law (M.S. 120B.30, 2004, subdivision 1). If only multiple choice items are included in the total test (i.e., no constructed response), testing time may not increase, even with the additional items.

If the intent of the federal legislation is to ensure that school AYP determinations are based solely on items covering content at the student's current grade level, and if Minnesota decided to use only multiple choice items (that is, to drop constructed response items), the test could begin with items administered adaptively, covering content in the standards at the students' current grade level (the grade level core)

followed by items administered adaptively, including some items from grade levels above or below the student's current grade level. School AYP status would be determined solely on items in the grade level core. The adaptive items at the end would give a more precise estimate of student achievement levels, particularly for students at the high and low ends of the achievement continuum. Students' overall scores would be based on both the adaptive and grade level core. If the grade level core could be administered adaptively, then the computerized adaptive test may require no more testing time than the current *MCA*s, even with constructed response items. Without constructed response items, it may take less time.

With the possible exception of constructed response items, the grade level core would be administered via computer. The difference between the two scenarios above is that in the first scenario, the grade level core is NOT adaptive, may include constructed response, and would include exactly the same items for all students. In the second scenario, the grade level core would be adaptive, and would not include constructed response items scored by human raters. Either way, the grade level core would be constructed of items that may be much the same as those in the planned *MCA-II*s. They would have to undergo the same review for content alignment, psychometric quality, and fairness as do the current *MCA* items. Even items in the fully adapted portion following the core would be much like multiple choice items in the *MCAs,* except that some of the items would correspond to content in grades above or below the student's current grade level. For these reasons, we refer to our recommendation as a conversion of the *MCA*s to computerized adaptive administration rather than calling it a replacement of the *MCAs*.

After completing the common core of items, testing would continue in a computerized adaptive fashion until the student's achievement level was measured to within the degree of precision specified in the testing process. Most students would also take some field test items that would not be included in their score. It is also possible that some items may be added to obtain more reliable subscale scores that can be compared to each other and to previous years. Scores reported to individual students would be based on all items administered (except field test items), not just those in the common core of items. If students receive a national percentile rank, that ranking would also be based on all items administered (excluding field test items), not just those in the common core.

Any diagnostic (subscale) scores would be based on all items administered (excluding field test items). Scores based on all (minus field test) items could be used to identify high performing schools, to assess growth for individual students and schools, and possibly to determine safe harbor or sanctions in the NCLB-mandated AYP evaluation of schools.

Typically, adaptive testing begins with an initial estimate of the student's ability and then proceeds to administer items adaptively until the student's score is estimated with sufficient precision. In the testing described above, where all students take the same items in the grade level core, a student's score on the grade level core would serve as the initial estimate of ability from which the adaptive testing would begin.

The proposal above that includes a common grade level core would yield an assessment and accountability process that should satisfy NCLB requirements. The proposal above that includes an adaptive grade level core may or may not satisfy NCLB requirements. Through conversations with USDE, MDE would need to clarify whether an adaptive grade level core would meet the federal requirements. This second option would seem to preclude inclusion of constructed response items scored by human raters, because it is unclear how one would combine results on constructed response items graded much later with the adaptive grade level core results to arrive at a determination of student proficiency for purposes of school evaluation in the AYP process.[2]

Embedding the Tests in a Larger Testing System

In order to connect results to some form of national norms and to diagnostic materials, this proposal includes augmenting the *MCA-II* items with items in an existing bank. NCLB guidelines permit such an augmentation so long as the resulting grade level core items are fully aligned with state grade level standards. In essence, the grade level core would contain a combination of items constructed specifically for Minnesota (essentially like our *MCA* items now) and items from an existing item bank of a test vendor. The goal is to construct a set of core items combining items from the existing bank and items written specifically for Minnesota so that the combination of items fulfills

[2] If the constructed items were electronically scored, rather than scored by human raters as is currently the case, constructed response items could be included in the adaptive portions of the test.

the test specifications and is fully aligned with grade level standards. Items in the adaptive portion would be drawn solely from the vendor's existing bank, although items may be selected from that bank to be aligned with Minnesota Academic Standards. In the fully adaptive portion, some items may be aligned with standards for grades above or below the grade in which the student is enrolled.

Constructing an augmented test begins by constructing a set of test specifications for each grade. The specifications include a list of the content areas to be covered, the total number of items in each content area, and the percentage (or number) of items in each content area. From the items on the existing tests (which may be a norm-referenced test, a diagnostic test, or both) one selects items that cover content listed in the test specifications. After having done so, there will almost certainly be some content areas that are not covered by items from the existing test. MDE and the vendor would write items to cover the missing content areas.[3]

To achieve the full potential of augmenting *MCA* items with items from an existing test, the items on the existing test must have been calibrated on a national sample, since national calibration connects results to national norms. The items need to be calibrated on scales for which there are existing interpretative and training materials useful for parents and teachers. Finally, the items need to be vertically equated so as to produce measures of student growth for schools.

On a cautionary note, national norms from a state-constructed test will not be fully comparable to those with which many are familiar. First, most norm-referenced tests have time limits. The normative information on the computerized, adaptive test would refer to untimed performance, whereas norm-referenced tests refer to performance in situations with a time limit. Second, not all national norms are based on a norm sample

[3] This process of constructing core items actually differs little from the current *MCA* process, because current vendors have a pool of existing items written previously for their various clients. Vendors will include some of those items, or minor adaptations of them, in an *MCA* if the item fits Minnesota's standards and test specifications. The major difference is not in the test construction process, but rather in the fact that the items of existing vendors have not been calibrated on a scale for which there are national norms; nor have they been calibrated on a vertical scale used to measure student growth from year to year; nor have they been calibrated on scales for which extensive diagnostic and interpretive materials are available. There is debate in the field as to whether one can construct and calibrate a single set of items that satisfy accountability, norming, and diagnostic purposes. While we agree that separate tests are better, separate tests also require more testing time. The diagnostic utility of statewide testing can be improved. Approximate national norm information can be provided, although the normative information will not be fully comparable to that provided by the best norm-referenced tests.

carefully constructed to be representative of the U.S. school population. Some vendors of computerized tests do not provide norms. It is possible that a computerized adaptive test vendor cannot provide normative information or that the vendor's norms are not as carefully constructed as those of the best norm-referenced tests. Even if norms are available, they will be based on untimed performance and therefore not fully comparable with more conventional norms based on timed performance.

Constructed Response Items

As previously discussed in connection with the writing tests, the teachers and administrators with whom we talked were skeptical that students were prepared to answer constructed response items on the computer. Since there is no statewide, grade specific standard as to when students should be able to compose text on computer, students cannot immediately be expected to do so. Furthermore, some current mathematics items require non-text answers (e.g., a chart or graph) and students do not necessarily know how to create such responses on computer. (See the MDE web site for examples of these types of questions www.education.state.mn.us/content.) Initially, if constructed response items are retained, they would have to be administered via paper/pencil and scored by human raters. Doing so has several disadvantages.

First, if constructed response items were administered by paper/pencil, the logistical convenience of computerized testing would be diminished. Answer sheets would have to be printed, delivered to schools, and returned to the vendors. MDE and schools would then have to manage two systems: a computer administered system and a paper/pencil system for every school.

Second, the costs would be substantially higher than for computer adaptive testing without constructed response items. Largely, these increased costs reflect human scoring of constructed response items plus the cost of the printing, delivery, and return of test booklets and documents. (See cost estimates on pages 34–37.)

Third, the constructed response items increase testing time. If only multiple choice questions were used, it may be possible to add the extra items for the adaptive portion without materially increasing testing time as compared to the current *MCAs*.

Fourth, constructed response items scored by human raters increase scoring time. These are the items that take the longest to score and pose the largest delay in results turnaround.

What would be lost if constructed response items were dropped? First, Minnesota law requires any test written after 2002–03 to include multiple choice and constructed response items (M.S. 120.B.30, 2004, subdivision 1). Furthermore, most educators agree that it is important to assess how students perform on higher order thinking skills and communication skills tapped by constructed response items. While, in our judgment, multiple choice items can tap many (if not all) higher order thinking skills, constructed response items tap the ability to organize and communicate knowledge in ways that multiple choice items cannot. If constructed response items are dropped, students and teachers may decide that the skills tapped solely by those items are unimportant and therefore those skills will receive less attention in instruction.

Despite the importance of constructed response items, the teachers and administrators to whom we talked had mixed reactions to dropping those items from the *MCA*. Teachers report that they learn a great deal about a student from reading the student's answer to constructed response items. Because *MCA* constructed response items are scored by the vendor and teachers do not read student's constructed responses on the *MCAs*, teachers reported that they did not learn as much from constructed response items on the *MCAs* as from constructed response items on classroom assessments. Further detracting from the importance of *MCA* constructed responses is the fact that teachers do not receive a separate score for students based on just constructed responses. All scores are based on a combination of multiple choice and constructed responses. Therefore the *MCAs* do not provide information about students' abilities to answer constructed items separate from their ability to answer the multiple choice questions.

When we asked people if they would favor dropping constructed response items if it would facilitate computerizing the tests, teachers responses were mixed; reactions of other groups were more favorable to dropping constructed response items. In the cost estimates section, we have included estimates with and without constructed response items. We recommend that Minnesota consider a change in Minnesota statue 120.B.30 to allow the Department of Education to drop constructed response items until such time as

students can reasonably be expected to enter responses on a computer keyboard. To hasten the day when students can enter responses on a keyboard, we recommend that Minnesota adopt standards that specify a grade when students can be expected to compose responses on a keyboard. When that standard is in place, writing tests and constructed response items could be administered by computer.

Length of reading passages

Delaying the inclusion of constructed response items is one option in our proposal. A second option involves the length of reading passages. As discussed in the literature review above, that literature identifies long reading passages with scrolling as a possible factor that may explain why students sometimes score lower on computerized reading tests. Comparability of scores is important for tracking trends over time and for fairness in any transition period in which some schools are taking the tests administered by computer while others are taking the tests in paper/pencil form. As discussed in the schedule below, there would be such a transition period.

We recommend that Minnesota consider the following options to help ensure comparability of reading passages administered paper/pencil and by computer:

- Shortening the reading passages so that they fit on a single screen.
- Employing a simulated "page turn" method rather than scrolling if long passages are retained.
- If scrolling or page turning is required, ensuring that the question remains on the screen as students scroll or page turn through the material.
- Allowing students to highlight key phrases if longer passages are employed.

Careful attention should be paid to line length and font size. Some consideration must be given to how students with low vision can be given large font text.

Selecting a vendor

Minnesota law requires that large contracts be awarded through an open competition which begins with the preparation of a request for proposals (RFP), proceeds through a review of submitted proposals, and ends with the awarding of a contract. If

Minnesota decides to implement computerized adaptive testing, some changes should be made in the structure of the RFP and the criteria for review. Some additional factors become considerations in the awarding of the contract: e.g., minimum hardware and software requirements, experience of the vendor with computer adaptive testing, and vendor ability to accommodate changes in computer administration (e.g., allowing highlighting of reading passages). In short, the specifications in the RFP must be revised.

Furthermore, if Minnesota is going to embed computerized testing in a larger assessment system for purposes of obtaining national norms and diagnostic information, then the RFP must ask vendors to describe their norms and norm group (if any), diagnostic scores, materials and scores that would help teachers and parents understand those scores, and any training available to assist in use of the information.

If the criteria listed in the RFP change and the information requested of vendors changes, then the composition of the review panel may also need to change. That is, the proposal review panel may need educational technology expertise to review minimum hardware/software requirements, classroom teacher expertise to review classroom score reports, diagnostic scores, interpretive materials, and training options. Choosing a vendor with computerized adaptive testing experience would also be necessary.

Implementation schedule

The following is a suggested schedule for implementation.

- 2005–06: Draw up and issue a request for proposals and select a vendor.
- 2006–07: Draft items in the spring and fall of 2006. Field test those items in 2007. Also, create sample computerized tests that districts can administer to students beginning in the spring of 2007. This will give students a chance to become acquainted with computerized testing, if they are not already. It will also give schools a chance to field test their technology.
- 2007–08: Begin administering computerized adaptive testing in spring 2008.

We envision a transition period in which schools will have a choice of test administration. During the transition period, most schools will administer the test via

computer but some may select paper/pencil. Some will as yet be unable to administer by computer. The transition period might last approximately three years. Depending on the minimum hardware/software specifications, we estimate (see the Feasibility section, p. 32) that up to 80% of schools could administer a computerized adaptive test with their current hardware configuration or with a modest alteration in the configuration if the testing window is lengthened to 4-5 weeks and constructed response items are administered by paper/pencil or dropped initially until students can be expected to compose on a computer. During the transition period, however, some schools capable of computerized testing may choose to administer paper/pencil rather than lose instructional time in the computer lab. Unfortunately, we do not have a good estimate of how many schools might make this decision.

Minnesota may decide to perform a study comparing results from paper/pencil and computerized testing before deciding whether to implement computerized testing. This would delay the schedule of implementation.

Possible Advantages

The biggest single advantage of computerized adaptive testing involves the choice between assessments covering basic content appropriate for all students vs. assessments covering more cognitively complex content or (above grade level) material studied by some, but not all, students. With a single paper/pencil test common to all students, one cannot create a single test of reasonable length which covers both. With an adaptive test, however, one can create an item bank that includes content from the most basic material covered by all students to the most advanced material reached only by some. The computer will adapt the item content to the achievement level of the student. Low achieving students will not be frustrated by items beyond their instructional level while high achieving students will have the opportunity to demonstrate their capacity to handle the most advanced work. Low achieving students may receive items covering content below their current grade level, if appropriate; while more advanced students may receive items covering content one or two grades above their current grade level.

The second advantage of computerized, adaptive testing follows from the first. For several years now, Minnesota legislation has called for the measurement of students'

grade-to-grade achievement gains and the use of those gains in a value-added evaluation of schools (M.S. 120B.30, 2004, subdivision 1a). If the statewide tests are adaptive, then student achievement levels can be measured more accurately (smaller standard errors of measurement). The biggest improvements occur for students at the low and high extremes of achievement. If scores are measured more precisely for these students, then year-to-year gains are measured more precisely. The *MCA-II* will provide growth scores, but computerized adaptive tests can provide more accurate growth scores. If year-to-year gains of students are measured more accurately, then value-added indicators of school quality, indicators based on student achievement gains, will be measured more accurately. Thus, computerized adaptive testing can improve a value-added system of evaluating schools.

Third, computerized testing (adaptive or not) would mean instantaneous results to students and teachers, at least results on multiple choice items. Classroom results on those items would be available as soon as the last student in the class completed the test and the same would be true for school level data. Students will then, hopefully, take the results more seriously. Scores returned weeks or months later are "old news."

Furthermore, quicker turnaround will facilitate use of the test results in school and district planning. Teachers and administrators commonly do planning for the following academic year during the summer. The planning must be based on data available early in the summer. If student scores on multiple choice items are available at the beginning of the summer, then teachers and administrators can use the data in the planning for the subsequent year. However, depending on details of the system, the instantaneous results to schools and districts may be only preliminary results based on multiple choice items. Final results would not be available until the constructed response items (if used) have been scored.

Fourth, if the computerized assessment is embedded in a larger testing system, that larger system may include diagnostic information useful to teachers and administrators. For instance, it may contain diagnostic subscales that have been carefully constructed to be of equal difficulty each year so schools can track improvements (or declines), not only in overall scores (e.g., mathematics) but also in more specific content areas (e.g., fractions, measurement). It may contain reading scores (e.g., lexile scores)

that are linked to reading levels of text materials and that can be used by teachers to plan the reading levels of reading instruction and leisure reading for students. The larger system may contain supplemental manuals that help teachers interpret the test scores diagnostically and plan interventions for students. And it may contain training that further helps teachers interpret the scores and plan further instruction. While computerized adaptive statewide tests would not eliminate the need for formative assessments, they could enhance the instructional utility of statewide summative assessments.

Finally, computerized adaptive testing should be less expensive to administer than paper/pencil tests. There are initial start up costs for setting up the system and providing the technology necessary to take the tests, and the start up costs will make the system more expensive than the current paper/pencil system during the transition period. However, once all schools have switched to computer administration, costs should be less than those associated with paper/pencil administration, because computerized testing has no printing, shipping, scanning or storing costs for test booklets, answer sheets or reports. Some of these cost savings may be offset by an increase in the need for technology equipment, support and maintenance, but there should still be some cost savings.

Possible Disadvantages

One possible disadvantage of the proposed computerized format is increased test length. Whether the test actually takes more student time, however, will depend on whether constructed response items are included and whether the grade level core is adaptive. Both elimination of constructed response items and adapting the core would shorten the test and, together, may lead to a test that is shorter in student testing time (possibly not fewer items) than is the current *MCA*. If, however, constructed response items are retained and the grade level core contained the same items for all students, students would take a grade level core of 40–50 items followed by items administered adaptively which would increase student testing time.

As stated previously, the exact number of items taken by students will vary from student to student in an adaptive test. However, if all students must take exactly the same items in the core, then some students may have to take approximately 50% more items than with the current *MCA,* while others may take no additional items, Since additional

items will all be multiple choice, testing time will not increase by as much as 50%. If Minnesota should decide to use only multiple choice items in the grade level core and adaptive items, the testing time may be approximately the same as with the current *MCA*s.

A second disadvantage to computerized testing is the lengthened testing window that would be required. The extended window would mean utilizing computers for testing during several weeks a year thereby reducing computer use for classroom instruction. The lengthened testing window may also lead to possible security concerns arising because students can talk about items with one another before some students have taken their tests. Finally, the lengthened testing window could create unequal opportunities to learn whether students in some schools or districts take the test weeks after other students. To minimize these concerns, we recommend that MDE require schools to test students in a given grade and subject within a two week testing window. To minimize security concerns, we recommend embedding the state test in a larger testing system to ensure that, due to larger item banks, students will be unlikely to see the same test items.

Given the minimum hardware and software requirements for administering multiple choice on computer, most of our schools have the technology to implement the testing now. However, this does not mean they have the technology to implement more sophisticated testing involving interactive items and complex graphics. While most schools have the technology to implement the testing envisioned here, they cannot do so without a significant loss of their technology to instruction.

Finally, there can be an effect of the delivery mode, computer vs. paper/pencil, on results. If the computerized tests are constructed carefully, there should be little or no effect. To the extent there is an effect, it can affect trends over time. During a transition period in which some schools are administering the test by computer while others are administering it paper/pencil, the effect can give a small advantage to one or the other set of students unless statistical adjustments are made to the data to compensate for known effects. Such an effect, if any, may be short lived as students become more accustomed to taking tests on computers.

Below is a table to help illustrate the advantages and disadvantages of the types of assessments discussed in this report:

Test Characteristics	Current *MCA* Paper/ Pencil Test	Proposed Computerized Fixed Test	Proposed Computerized Adaptive Test
Allows for item review	X	X	
Fixed length	X	X	
Fast turnaround time for results (multiple choice or electronically scored constructed response)		X	X
All students receive the exact same items	X	X	
Precise measurement of year-to-year student growth			X
Questions matched to student ability			X
Potential to include diagnostic information	X	X	X
Potential to include national norms			X
No limit to the number of students tested at one time	X		
Limited printing or shipping costs		X	X
Increase in technology costs		X	X

Feasibility

In order to gauge the current technology capacity of schools in Minnesota, we conducted a technology survey across the state. The survey was developed by the Office of Educational Accountability in collaboration with an advisory committee. It consisted of 29 questions and was administered online. The questions covered the number and types of computers that schools have for student use; computer networking; internet connections; school and district server capacity, and technology staff (See Appendix E for a copy of the survey).

Virtually all respondents (99%) reported their school has networked computers available for student use in a lab or media center setting. Sixty-eight percent have two or more student computer labs in their school. Nearly all (99%) said that student computers are connected to either a school or district server and all are connected to the Internet.

To estimate what percentage of Minnesota schools could administer computerized tests with the systems currently in place, responses were analyzed based on the minimum technology requirements of two vendors. In order for a school to be classified as capable, its technology system had to have the following:

- networked computers for student use in a computer lab or media center;[4]
- computers connected to a school OR district server;
- student workstations connected to the Internet directly or via a school or district server;
- 6 kilobytes of sustained bandwidth available per student computer;
- student workstations with at least Pentium (100 MHz) processors (for PCs); MAC or iMAC 233 MHz (for Macs)
- student workstations with at least 32 MB RAM;
- student workstations have an operating system of at least Windows 98 (for PCs); MAC OS 7.5 or higher (for Macs); and
- student workstations have at least 640 x 480 VGA monitor resolution.

Based on these criteria, approximately 70% of the schools that answered our survey currently have a technology infrastructure that would allow them to do the adaptive computerized testing discussed in this report. Another 10% could do so with a relatively inexpensive addition of a caching station to the computer workstation (estimated average cost $800 per computer lab). That is, approximately 10% of the schools met all of our criteria except for the availability of bandwidth. Storing student responses on a caching station (essentially a desktop computer wired to student workstations) until bandwidth was available would allow an additional 10% of schools to administer the computerized adaptive tests envisioned here. Thus, almost 80% of schools seem to have the necessary hardware/software capacity or could acquire it with a relatively inexpensive addition to their systems.

---

[4] Any schools that said they had student computers but not in a lab setting were classified as not having networked computers. For the purpose of high stakes statewide testing it is crucial to be able to test an entire class in one room at the same time. At schools where student computers are scattered throughout the building this poses security and supervisory concerns.

As mentioned previously in this report, this could be an underestimate due to the wording of questions on the survey relating to computer and server specifications. Some respondents answered that their school or district systems do not meet certain specifications but they actually meet higher specifications. This means that they may not have been classified as capable when truly, they are more than capable of successfully administering these tests.

This estimate could also prove limited because, although some schools were classified as having the technology capacity currently, it is unclear whether or not all of their systems will meet the necessary specifications by the time the testing is implemented. Some of the workstations may be out of date in several years or server capacity or availability could change. It is also possible that vendors may increase minimum requirements by the date of implementation.

Cost

Here we summarize our overall cost estimates for computerized adaptive testing and compare those to costs of the *MCA* testing program in 2006. The cost of computerized adaptive testing can be divided into two categories: (1) ongoing administration and operational costs; and (2) capital costs of workstations, software, and networking. Both are ongoing costs, but operational costs are an annual cost, whereas the capital costs of updating infrastructure would occur on a 4 - 6 year basis. A third possible cost is the expense of a pilot study comparing student results on computerized and paper/pencil tests. Some states, (e.g., Kansas), have undertaken such a study before implementation of computerized test. The estimated cost of such a study is included in Appendix F, although the precise cost would depend on the extensiveness of the desired study.

Annual Operational Costs of Test Administration

The first category can be compared to what the state currently spends on statewide assessments. For the purpose of this report we are looking at the cost of reading and mathematics assessments in grades 3–8, reading in grade 10, and mathematics in grade 11. Based on 2005–06 estimates, the cost to continue the *MCAs*—or *MCA-IIs*—for

grades 3 through 8 and grades 10 and 11 in reading and mathematics would be approximately $8 million (see Appendix G). This estimate assumes testing eight grades of students with approximately 63,000 students in each grade. These costs include the following elements of an assessment system: *item development, printing test materials, providing support materials, distribution and collection of test booklets and materials, test security, processing and scoring tests, reporting test results, technical reporting and review, quality assurance, timelines and scheduling, and training*. This estimate does not include the following additional costs: *science test development, state and federal staff, teacher review/OEA/experts, 5th grade writing test*, and the *10th grade writing test.*

We spoke with two vendors to acquire estimates on what the test administration portion of this system would cost using a computerized format. If constructed response items scored by human raters were excluded from the computerized adaptive test, the costs that were quoted would be very comparable to what the state would spend to continue paper/pencil tests in its current form (with constructed response). The vendor estimates included the same components of the assessment system mentioned in the above paragraph only with computer administration rather than paper/pencil administration and electronic files rather than paper booklets and reports. In addition, schools would have the option to administer the assessments at least one other time during the school year, but up to as many as three additional times depending on the vendor. This price also includes support/interpretive materials for teachers in using the results for instructional purposes.

These estimates were compiled assuming that, at least in the beginning, there would be a percentage of students or schools administering the tests via paper/pencil while others were using computer delivery. We assumed a 70/30 split with 70% of schools administering via computer and 30% via paper/pencil. Therefore, in some regards, the state would be paying to administer a dual system for a transition period.

One vendor provided per-student costs for computer assessment administration and paper/pencil administration. For the schools taking the test via computer, the cost would be $8.50 for computerized adaptive test administration plus $12.50 for constructed response items per pupil per subject. For schools administering the tests via paper/pencil, the cost would be $7.50 plus $12.50 for constructed response per pupil per subject tested.

The per-pupil cost for paper/pencil tests is not really less than those for computerized administration, because this vendor "loaded" many of the fixed costs (e.g., item development, score reporting) onto the computer administration estimate.

If Minnesota chooses to test grades 3–8, 10, and 11 in reading and mathematics, assuming 63,000 students per grade, then the overall operational cost in reading and mathematics (excluding any constructed response items) would be $8–9 million. If the state decides to continue with constructed response items in the current form the cost would be an additional $12.6 million for testing in reading and mathematics. Thus, depending on whether constructed response items are included, the cost could range from $8 million to $21.3 million. (See Appendix G).

The second vendor provided an overall estimate of $5–$10 per pupil per test for a 70/30 split between computer administration and paper/pencil administration. Therefore, in calculating an estimate we took a conservative estimate of $8.50 per pupil per test or $17.00 per student for reading and mathematics or about $9 million. These costs include all of the administrative aspects of the current system (excluding constructed response items). If the state includes constructed response items, the cost may increase to approximately $22 million, largely depending on whether the constructed response items included were scored electronically or by human hand-scoring. (See Appendix G).

Capital Costs, Workstations, Software, and Networking

It is difficult to estimate the cost of getting districts that are not currently equipped to administer computer based tests properly equipped to do so. In some cases, where the district needs only to add caching stations or upgrade a server, the costs may be minimal (i.e., less than $2,000 per school).

Although it is fairly easy to estimate the cost of student workstations, it is not at all easy to estimate the cost of networking a school or district. These costs are influenced by factors such as how the school is structured, how far individual buildings are from each other and the district office, and the distance from metropolitan areas.

In some cases, schools cannot add any more separate labs because they lack additional classrooms. The only way to add capacity in this situation would be to purchase portable laptops that can be used in an existing classroom and install wireless

hubs. Suitable laptops are also usually more expensive than desktops. The additional wiring of such schools is extremely difficult to estimate since the cost would depend on the particular circumstances of the school.

We asked two districts how they estimated their costs in their technology plan. Both provided an estimate for computer lab costs which included a workstation (computer hardware and software) and networking (including servers) component. The average of these two district estimates was $30,000–$35,000 per lab of 30 computers. Based on these estimates, we anticipate that districts will need to spend approximately $25,000–$35,000 per lab every 4–6 years in order to maintain and upgrade the hardware/software necessary to conduct the kinds of computerized testing envisioned in this report. Costs would depend on whether schools had to simply replace computers or also had to upgrade networking. Such facilities would, however, not be solely dedicated to testing. We would anticipate that they would be dedicated primarily to instruction except during the spring testing window.

Conclusions/Recommendations

In this section, we summarize the conclusions and recommendations made above.

General Guidelines

- Care must be taken to ensure comparability of computerized and paper/pencil forms of the test. Comparability of computer administered and paper/pencil results will be important for purposes of tracking trends over time and for fairness to students during a transition in which some students may still be taking paper forms of tests.

- Currently the testing window for statewide tests is three weeks. That is, schools have three weeks in which to administer the tests. Given the technology available in schools, the testing window would need to be lengthened to allow approximately four to five weeks for students to complete all exams in all grades. In most schools, one exam (i.e., one subject in one grade) could be completed in a single day. We recommend that MDE require schools to test students in a given subject and grade within a two week testing window, but multiple subjects and multiple grades plus make-ups would extend over four to five weeks depending on the number of students being tested and the number of networked computers available for testing. It is worth noting that as more tests (e.g., science tests) become computerized the testing window will lengthen accordingly.

Adaptive Testing

- We are recommending that computerized adaptive testing begin with reading and mathematics in grades 3–8, reading in grade 10, and mathematics in grade 11. We are not recommending computerizing the current *MCA* writing assessments until students are proficient at keyboarding and can construct their answers on the computer rather than via paper/pencil.

- To satisfy the testing requirement of NCLB, we propose that the adaptive test begin with on grade level items aligned with the state standards administered in a fixed or adaptive format as permitted by the U.S. Department of Education (USDE). The on grade level portion of the test would be followed by a fully adaptive portion. In the adaptive portion, each student would take a unique set of items, but within that unique set of items, there would be a subset covering content from the Minnesota standards at the student's current grade level. For the purposes of the federal accountability system, schools would be evaluated on how well their students performed on grade level items only.

- If Minnesota would like the grade level core to be adaptive and if Minnesota is willing to drop constructed response items, it would need to open discussions with USDE as to whether it would allow AYP to be determined by an adaptive test composed solely of multiple choice items aligned with grade level standards. If the core were also adaptive, testing time would be reduced with no loss of accuracy in student scores.

- If Minnesota adopts computerized adaptive testing, then it will need to revise the information required of vendors through the request for proposals process, the criteria used to review the proposals, and the expertise represented on review panels. The criteria should cover such things as normative information, diagnostic materials, score reports including online reports available to teachers and provisions for growth scores, hardware/software requirements, experience with large-scale computerized adaptive assessment, and training provided for interpretation of results. The expertise represented in review panels should include teachers and educational IT people for purposes of evaluating reporting, training, and software/hardware requirements.

Reading

- In developing reading tests, Minnesota should design reading items so that students can scroll through a passage while retaining the item in view. Passages should use lines of intermediate lengths, roughly 55 characters per line for older students and possibly shorter lengths for younger students. Since the literature is unclear as to whether scrolling adds to the difficulty of reading, Minnesota should consider the following options:
    - Shortening reading passages such that the entire passage and items can fit on a single screen.

- Employing a simulated "page turning" technology in which students do not scroll through a passage, but rather click a button to turn to the next page while retaining the item in view.
- Allowing students to highlight key phrases if longer passages are employed.

- Careful attention should be paid to line length and font size. Some consideration must be given to how students with low vision can be given large font text.

Writing

- To set the stage for computerized writing assessments, we recommend that Minnesota establish a content standard stating a grade by which students are expected to compose written material on a keyboard utilizing spell checking and grammar checking to remove spelling and grammatical errors. Once these standards are in place, writing assessments should be converted to computer administration.

Constructed Response

- Minnesota may decide to keep constructed response items along with a computerized adaptive test. If, however, Minnesota should decide to delay inclusion of constructed response items on computerized adaptive tests, rather than drop constructed response items, the state should take steps to prepare for the later inclusion of computer administered constructed response items. It can do so in two ways. First, as stated above, it should amend the Minnesota Academic Standards to state an expected grade at which students are expected to compose text answers on computer. Second, it should develop non-text/graphical constructed response formats that can be used to tap higher order thinking skills and communication skills in mathematics. Students will need to learn to respond in those formats.

- If Minnesota adopts a computerized adaptive testing format, for reasons of cost, testing time, and logistical feasibility, the state should consider using only multiple choice items initially and delaying inclusion of constructed response items until students can enter their responses to such items on computer.

- Computerized testing is coming to statewide testing in Minnesota. In the tests that many districts administer to their students, computerized testing is already here. It remains to be decided when and how Minnesota will make the transition to computerized statewide testing and whether the computerized assessments will be adaptive.

References

Dyson, M.C. & Haselgrove, M. (2001). The influence of reading speed and line length on the effectiveness of reading from screen. *International Journal of Human-Computer Studies, 54*, 585-612.

Gambrell, L.B., Bradley, V.N., & McLaughlin, E.M. (1987). Young children's comprehension and recall of computer screen displayed text. *Journal of Research in Reading, 10*(2), 157-163.

Omnibus Education K-12 Education Policy Bill 2004, H.F. 1793, 83d leg, (2003-2004).

Kiely, G.L., Zara, A.R., & Weiss, D.J. (1986). Equivalence of computer and paper-and-pencil Armed Services Vocational Aptitude Battery Tests (AFHRL-TP-86-13). Brooks Air Force Base, TS: Air Force Human Resources Laboratory.

Lunz, M.E., Bergstrom, B.A., & Wright, B.D. (1992). The effect of review on student ability and test efficiency for computerized adaptive tests. *Applied Psychological Measurement, 16*(1), 33-40.

Mazzeo, J., & Harvey, A.L. (1988). The equivalence of scores from automated and conventional educational and psychological tests: A review of the literature. College Board Report No. 88-8. ETS RR No. 88-21.

Mead, A.D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin 114*(3), pp. 449-458.

Meyer, J.F. & Poon, L.W., (1997). Age differences in efficiency of reading comprehension from printed vs. computer-displayed text. *Educational Gerontology, 23*, 789-807.

Minnesota Statute § 120B.30 (Office of Revisor of Statutes 2004). Retrieved June 3, 2005 from http://www.revisor.leg.state.mn.us/bin/getpub.php?pubtype=STAT_CHAP&year=current&chapter=120B

The No Child Left Behind Act of 2001, 20 U.S.C.A. § 6301 et seq. (West 2002).

Poggio, J., Glasnapp, D., Yang, X., & Poggio, A. (2004). An experimental study of score results from computerized and paper-and-pencil mathematics testing in a state large scale assessment program. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA, April 2004.

Reinking, D. (1988). Computer-mediated text and comprehension differences: The role of reading time, reader preference, and estimation of learning. *Reading Research Quarterly, 23*(4), 484-498.

Rice, G.E. (1994). Examining constructs in reading comprehension using two presentation modes: Paper vs. Computer. *Journal of Educational Computing Research, 11*(2) 153-178.

Standards and Assessments Peer Review Guidance: Information and Examples for Meeting Requirements of the No Child Left Behind Act of 2001 (2004). Retrieved June 6, 2005 from www.ed.gov/policy/elsec/guid/saaprguidance.pdf

Stone, G.E., & Lunz, M.E. (1994). The effect of review on the psychometric characteristics of computerized adaptive tests. *Applied Measurement in Education 7*(3), pp. 211-222.

Title I Directors Conference (February 2003). Retrieved June 3, 2005 from

www.ed.gov/admins/lead/account/standassess03/edlite-slide17.html,

www.ed.gov/admins/lead/account/standassess03/edlite-slide23.html

Vansickle, T.R.V. (2005). Minnesota Assessment System. Presentation to the Minnesota Elementary and Secondary Principals Association Institute, 2005.

Appendix A: Personal Interview Questions

Computer Assessment Project
Personal Interview

Name: _____

Date: _____

1. Does DISTRICT use the computerized or paper/pencil version of the NWEA?

2. If paper/pencil, have you tried the computerized version? What was your experience?

3. What do you see as advantages and disadvantages of the NWEA?

   Turn around time?

   Administering the test?

   Reporting? (teacher and parent)

   Cost?

   Measuring Student Growth?

4. What is the current infrastructure in DISTRICT? T1 or DSL lines? Computer lab size? Type of machines? Browser? Java script?

5. Would you be in favor of the NWEA if the content was significantly altered to meet NCLB requirements?

6. Regarding content: Which grades is it appropriate? Is it aligned with standards? What is the difficulty level compared with the MCAs/standards? What is your feeling about an all multiple choice test?

7. Does NWEA currently provide you with any web-based resources? (primarily for reporting purposes?)

8. Do you have any security concerns? How is it currently handled?

9. How is the customer service with NWEA? Technical expertise?

10. Suggestions for moving forward with this project?

11. Survey of technology?

Appendix B: Vendor Interview Questions

1. What computerized or computerized adaptive testing products does VENDOR have now, and what products will you have in the near future?

2. Please describe item development, ordering of tests by schools, test delivery to schools, test administration in the schools, uploading of student results, and reporting of results.

3. What are the hardware and software requirements in the schools and in the state agency for use of those products?

4. In rough figures, what would be the cost of a statewide testing program using the product(s) as a platform?

5. Computerized testing requires some retraining of staff in schools and the state agency. How does this take place with your product?

6. Can you give us examples of districts or states that are using the product? What has been their experience? Can you give us a name of someone to whom we can talk in that district or state agency?

Appendix C: Focus Group Questions

Computerized Assessment Focus Group Questions

1. In addition to the MCA's what other tests does your district currently administer?

2. Do you administer any computerized tests in your district?

3. How instructionally informative are the current MCAs?

4. What, if anything, would make them more instructionally useful? (turn around time, reporting, spring vs fall, growth measures)

5. How about the NR tests administered in your district? How instructionally informative are they? What, if anything, would make them more useful?

6. What is the current cost of the NR testing in your district?

7. What is the current infrastructure in your district? T1 or DSL lines? Computer lab size? Browsers? Types of machines?

8. How would you feel about switching from current statewide tests to computerized statewide tests?

9. What if switching meant eliminating CR items? Quicker turn around time? Giving up instruction time in computer labs?

10. What security concerns to have regarding testing in your district?

11. Names of people to talk to in the district?

Appendix D: Contacts

Computerized Assessment Project
August 15, 2004 – June 15, 2005
Contacts

Minnesota Department of Education:
Alice Seagren
Tim Vansickle
Pat Olson

Vendors:
NWEA
Riverside
Pearson
ETC
Scantron

Districts:
Minneapolis
St Paul
Bloomington
Rochester
Duluth
Lakeville

Legislature:
Steve Kelly
Gen Olson
Barb Sykora
Mindy Greiling
Lisa Larsen
Cap O'Rourke
Brent Gustufson
Eric Nauman
Mike Roelofs

Advisory Committee:
Jim Angermeyr (Bloomington)
Dave Weiss (U of MN)
Tim Hodges (Farmington)
Bob Kochmann (Sauk Rapids)
Dirk Mattson (MDE)

Education Associations:
Minnesota Association of School Principals
Minnesota Association of Elem Principals
Minnesota Association of Sec. Principals
Education Minnesota
Parents United
Association of Metro School Districts
Minnesota Rural Education Association
Minnesota School Boards Association
Minnesota Association of School Curriculum Directors
Minnesota Assessment Group
Non-public School Council
National TAC
Local TAC
Stakeholders Group

Appendix E: Zoomerang Online Survey

School and District Technology Survey

1

District Name

2

District Number

3

School Name

4

School Number

5

Contact Name

6

What grades do you have in the school? (Please check all that apply)

- [ ] Kindergarten
- [ ] Grade 1
- [ ] Grade 2
- [ ] Grade 3

- [ ] Grade 4
- [ ] Grade 5
- [ ] Grade 6
- [ ] Grade 7
- [ ] Grade 8
- [ ] Grade 9
- [ ] Grade 10
- [ ] Grade 11
- [ ] Grade 12

7

Does your school have networked computers available for student use in a lab or media center setting?

YES NO

8

How many labs does your school have? (Please include mobile/wireless labs)

- [ ] 1
- [ ] 2
- [ ] 3
- [ ] 4
- [ ] more than 4

9

How many total networked computers does your SCHOOL have available for student use in a lab setting?

10

How many networked computers does your DISTRICT have in student computer labs?

11

Do your computer labs have a teacher workstation that has the capability to send information to the student workstations as well as receive information from the school or district server?

YES NO

12

To what type of server(s) are the student computers in the labs connected? If the lab is connected to a school server AND a district server, please check "Other" and write in "school server and district server."

- school server in a LAN
- school server in a WAN
- district server in a LAN
- district server in a WAN
- none of the the above
- Other (please specify)

13

How are student workstations in computer labs at the school connected to the Internet?

- Not connected to the Internet
- Connected directly to the Internet using Microsoft Internet Explorer 5.01 or higher; or Netscape Navigator 6.1 or higher; or Apple Safari 1.0 or higher

- Connected to the Internet via school server using Microsoft Internet Explorer 5.01 or higher; or Netscape Navigator 6.1 or higher; or Apple Safari 1.0 or higher
- Connected to the Internet via district server using Microsoft Internet Explorer 5.01 or higher; or Netscape Navigator 6.1 or higher; or Apple Safari 1.0 or higher

14

What is the sustained bandwidth of the network connecting the student stations in the computer labs to the Internet?

15

What is the peak bandwidth of the network connecting the student stations in the computer labs to the Internet?

16

What types of student computers do you have in your school computer labs?

- Macintosh
- PC
- Both

SUBMIT

Survey Page 1

School and District Technology Survey

17

Which of the following specifications does your school server meet?

(Please check all that apply).

- [ ] Pentium II (266 MHz)
- [ ] MAC G3 or MAC G4
- [ ] At least 128 MB RAM
- [ ] At least 250 MB available disk space
- [ ] Operating system of Windows NT Server 4.0, SP4; or Novell 4.0; or Windows 2000 Server; or Windows 2000 Advanced Server; or Windows 2003 Server
- [ ] Operating system of MAC OSX Server
- [ ] Network of Windows NT or Novell 4.0
- [ ] Network of MAC using AFP
- [ ] No school server

18

Which of the following specifications does your district server meet? (Please check all that apply)

- [ ] Pentium II (266 MHz)
- [ ] MAC G3 or MAC G4
- [ ] At least 128 MB RAM
- [ ] At least 250 MB available disk space
- [ ] Operating system of Windows NT Server 4.0, SP4; or Novell 4.0; or Windows 2000 Server; or Windows 2000 Advanced Server; or Windows 2003 Server
- [ ] Operating system of MAC OSX Server
- [ ] Network of Windows NT or Novell 4.0
- [ ] Network of MAC using AFP

- [ ] No district server

19

Which of the following sets of specifications do your student workstations meet? (Please check all that apply)

- [ ] Pentium (100MHz) processor or higher
- [ ] Pentium II (266 MHz) processor or higher
- [ ] 32 MB RAM
- [ ] 128 MB RAM
- [ ] 500 MB available disk space
- [ ] Operating System of Windows 98 or above
- [ ] 800x600 SVGA Monitor Resolution
- [ ] 640 x 480 VGA Monitor Resolution

SUBMIT

School and District Technology Survey

20

How many of the networked Macintosh computers in your student computer labs are iBook, PowerBook, or iMac?

- [ ] We have no Mac computers in student labs
- [ ] All
- [ ] More than 75%
- [ ] 50%- 75%

- [ ] 25% - 49%
- [ ] Fewer than 25%
- [ ] None

21

Which of the following specifications does your school server meet? (Please check all that apply.)

- [ ] Pentium II (266 MHz)
- [ ] MAC G3 or MAC G4
- [ ] At least 128 MB RAM
- [ ] At least 250 MB available disk space
- [ ] Operating system of Windows NT Server 4.0, SP4; or Novell 4.0; or Windows 2000 Server; or Windows 2000 Advanced Server; or Windows 2003 Server
- [ ] Operating system of MAC OSX Server
- [ ] Network of Windows NT or Novell 4.0
- [ ] Network of MAC using AFP
- [ ] No school server

22

Which of the following specifications does your district server meet? (Please check all that apply.)

- [ ] Pentium II (266 MHz)
- [ ] MAC G3 or MAC G4
- [ ] At least 128 MB RAM

At least 250 MB available disk space

- [ ]
- [ ] Operating system of Windows NT Server 4.0, SP4; or Novell 4.0; or Windows 2000 Server; or Windows 2000 Advanced Server; or Windows 2003 Server
- [ ] Operating system of MAC OSX Server
- [ ] Network of Windows NT or Novell 4.0
- [ ] Network of MAC using AFP
- [ ] No district server

23

Which of the following specifications do your networked student workstations in computer labs meet? (Please check all that apply.)

- [ ] MAC (Power PC Based only)
- [ ] iMAC 233 MHz
- [ ] 32 MB RAM
- [ ] 128 MB RAM
- [ ] Operating system of MAC OS 7.5 or higher
- [ ] Network of MAC using AFP
- [ ] Monitor resolution of 800 x 600 VGA
- [ ] Monitor resolution of 640 x 480

SUBMIT

School and District Technology Survey

24

What is/are the title(s) of the staff member(s) at your SCHOOL who is/are responsible computers? (Please check all that apply)

- [ ] Technology Director
- [ ] IT Director
- [ ] Information Systems Manager
- [ ] Media Specialist
- [ ] Media Paraprofessional
- [ ] Other (please specify)

25

On a scale of 1-5, with 1 being not at all feasible and 5 being very feasible, how feasible would it be for your school to devote two hours of time per student within a four week window to give a computerized test? The total amount of testing hours would be two times the number of students divided by the number of networked computers in computer labs.

not at all feasible				very feasible
1	2	3	4	5

26

What, if anything, would make it more feasible?

27

Is Macromedia's Flash Player currently installed on computers in your student lab(s)?

YES NO

28

Do you have headphones for all computers in the lab that allow students to listen to audio media?

YES NO

29

What is your technology replacement cycle for student computer labs?

- less than 3 years
- 3-4 years
- 5-6 years
- 7-8 years
- more than 8 years
- Other (please specify)

SUBMIT

Appendix F: Comparability Study

Purpose

This study is designed to provide information on the comparability of scores derived from paper-and-pencil (P&P) and PC-based computerized administration of subtests of the *Minnesota Comprehensive Assessments* (*MCAs*).

Design

The study will use an alternate forms retest design. Students who have taken the standard Reading and Mathematics *MCAs* by P&P will be retested after a specified time interval. Students will be randomly assigned to either P&P or PC-based alternate forms of the same tests. On retest, all students will take both a reading and a math test using the same mode of administration. This design will allow for analysis of the following psychometric characteristics of the tests administered under the two modes:

1. Item statistics and internal consistency reliabilities.
2. Score distributions.
3. Alternate forms retest correlations.
4. "Construct validity" correlations between scores on the two tests for each of the two modes of administration.

Students

To examine the effects of mode of administration (P&P versus PC) on different types of students, samples of students will be selected from Grades 5, 8, and 11 in urban, suburban, and rural school environments. For each of these 9 cells of this design, 150 students will be randomly selected for each experimental condition (P&P and PC) for a total of 2,700 students. To attain the 150 students for each of the nine cells of the design, groups of 25 students will be randomly selected from classes at the specified grade level per classroom and randomly assigned to the P&P or PC conditions. This process will be repeated in additional schools as required until 150 students have been tested in each cell of the design. Each student will be administered the experimental tests (both P&P and PC) under controlled and carefully supervised conditions in school computer laboratories during a single testing session of about 1.5 hours. Their scores on previously administered *MCAs* will be retrieved from school files and matched with those of the experimental tests for analysis.

Additional data available on each student will be retrieved from their records, including gender, race, eligibility for free/reduced lunch, English proficiency, and special education status.. These variables will be analyzed in relation to the test scores under both modes of administration to further evaluate the comparability of scores from the two administration modes. In addition, following completion of the PC-based tests, students will be asked to evaluate their test-taking experience on a standard set of rating scales.

Tests

The reading tests in the *MCA*s have relatively long passages, with a number of questions asked about each passage. To implement these tests on a PC, two different conditions will be created. In both conditions, a split-screen will be used with the reading selection on the left side of the screen and the questions on the right side. In addition, a "highlighter" will be available to students to allow them to highlight any portions of the passages that they desire. Students will be randomly assigned to one of two conditions:

1. *The scrolling condition.* In this condition, the left side of the screen will be scrollable, so that the student can scroll the screen reading passage up or down at will to view portions of the selection. The right side of the screen will remain stationary. Students will be required by the test administration software to demonstrate that they know how to scroll before the test will begin.
2. *The page-turning condition.* In this condition, one page of information will be displayed in the left-hand screen partition at a time. To simulate the kind of page-turning that occurs in paper-and-pencil testing, buttons at the bottom of that partition will allow the student to display the next or previous page of the selection. Before the test will begin, the student will be given the opportunity to practice "turning pages."

The math tests will use only the multiple-choice items in the *MCA*s. For analysis purposes, the original *MCA*s given to the students in this study will be rescored using only the item responses from the multiple-choice items.

All tests will be untimed, but testing time will be recorded for each test under both conditions to allow comparisons of testing time between modes. To maximize the similarity between the P&P and PC conditions, students will be allowed to re-visit (review) items within a test under the PC condition. That is, as in a P&P test, students will be able to go back to previous items at any time during the test. When they reach the end of the test they will be notified and will have the option to continue reviewing items or to terminate the test. For the reading tests, this type of review will be restricted to the questions about a single reading passage.

Data Collection

All data collection will be supervised by project staff to insure comparability of administration conditions. Project staff will select students for participation, contact schools and teachers to make arrangements, and supervise the administration of all tests. A larger number of students will initially be selected so that if students do not report for testing when scheduled, additional students will be scheduled for a later testing session. Students will be tested in groups, with the group size dependent on facilities available at each school.

Assuming that the standard P&P *MCA*s are administered in mid-February, data collection with the experimental tests will be scheduled to begin in early March 2006 and continue into May. The period from September 2005 will be used to prepare the experimental tests and pretest them on small groups of students, identify schools to participate, select students from file data, install PC tests on lab computers in the schools, and schedule students for testing. File data on the standard *MCA*s will be obtained after they have been completed and scored for students who have participated in the research. Data analysis will be completed during the summer of 2006, with a final report to be delivered in September 2006.

Budget

The project will require two 50% time Graduate Research Assistants to make arrangements for testing, prepare the tests both for P&P and PC administration, supervise all testing, retrieve data on the standard *MCA*s, analyze the data, and draft the final report. The project will be supervised by Professors Mark Davison and David J. Weiss of the University of Minnesota. PC-based testing software will be made available at no cost by Assessment Systems Corporation of St. Paul MN.

Graduate Research Assistants: 2 @ 50%. 12 months	$33,800
Travel and per diem for data collection	7,500
Software modifications for PC-based testing (if required)	3,500
Printing of P&P test booklets	1,000
Scannable answer sheets	500
Scanning charges	250
Long-distance in-state telephone charges	500
Miscellaneous office and computer supplies	750
Total	**$47,800**

Appendix G: Cost Estimates

Assumptions

Our analysis assumes the federal government approves the assessment program. If they do not, then the state would jeopardize No Child Left Behind (NCLB) funds for non-compliance totaling over $200 million for Fiscal Year 2005 – 06.

The computerized assessment costs included are based on reasonable assumptions and estimates provided by two vendors. The estimates for the paper/pencil MCA's were prepared by the Minnesota Department of Education. Vendor 1 estimates are $17.00 per pupil for reading, mathematics, and language usage computerized assessments. Science tests are an additional $2.00 per pupil. These costs include multiple choice and short answer or fill in the blank questions that will be scored electronically. These estimates do not include constructed response items that require human scoring. If those items were included, the cost would be an additional $25.00 per pupil. These estimates include four testing options per year per pupil. These costs are based on the assumption that at least one-half the students enrolled in the state are tested each year.

The cost estimates are $15.00 per pupil for reading, mathematics, and language usage paper/pencil assessments for Vendor 1. These estimates include a fall and spring testing for each pupil. Our total estimates for Vendor 1 are based on 70% taking the tests via computer and 30% taking the tests via paper/pencil.

Vendor 2 estimates are $5.00 - $10.00 per pupil per test. This vendor was not able to provide more detailed estimates without specific details. Therefore in calculating the estimate we took a conservative estimate of $8.50 per pupil per test or $17.00 per student for reading and mathematics combined. These estimates are only for computer administrated tests, not paper/pencil tests.

An unknown factor is the extent the infrastructure is in place in Minnesota to accommodate computerized testing for all students within a reasonable testing window. Our estimate, based on surveying schools, is that 70% of the schools have the infrastructure in place and an additional 10% could be ready with very minor upgrades. Some schools will need to add computer labs at an approximate cost of $30,000 - $35,000 per lab. These costs include 30 workstations, 30 "drops" to connect the workstations to a server, hubs, panels, switches, routers and wiring to connect the server to the internet. Costs of any DSL, cable, T1 lines, fiber optic cables etc. would be in addition, and we were unable to estimate these additional costs because the costs would vary so widely depending on such factors as the location of the school.

The number of estimated students to be tested is based on a K-12 enrollment of 825,000 students. (Approximately 63,000 per grade.) The number of students tested would be 504,000 (grades 3-8, 10 and 11). Two tests (reading and math) would be given in grades 3 – 8 and one test in grade 10 (reading) and one in grade 11 (math).

It is assumed that the efforts of state and federally funded assessment staff will be redirected to implement the new computerized assessments with no further staffing required.

Testing in science at three grades will begin in 2005 with development, 2006 includes field testing and in 2007 the test will be piloted in preparation for full implementation in 2008.

MCAs	FY 2006 Costs	Vendor 1 Costs	Vendor 2 Costs
Reading – 3	$542,379	$590,400	$612,000
Math – 3	$533,568	$590,400	$612,000
Reading – 4	$531,615	$590,400	$612,000
Math – 4	$523,821	$590,400	$612,000
Reading – 5	$529,966	$590,400	$612,000
Math – 5	$522,328	$590,400	$612,000
Reading – 6	$526,259	$590,400	$612,000
Math – 6	$518,972	$590,400	$612,000
Reading – 7	$530,475	$590,400	$612,000
Math – 7	$522,789	$590,400	$612,000
Reading – 8	$532,424	$590,400	$612,000
Math – 8	$524,554	$590,400	$612,000
Reading – 10	$596,553	$590,400	$612,000
Math – 11	$592,707	$590,400	$612,000
Subtotal	**$7,528,410**	**$8,265,600**	**$8,568,000**
Quality Assurance	$32,200	$32,200	$32,200
Timeline and Scheduling	$38,500	$38,500	$38,500
Technical Reporting and Review	$350,000	$350,000	$350,000
TOTAL	**$7,949,110**	**$8,686,300**	**$8,988,700**
Total with Constructed Response	**$7,949,100**	**$21,286,300**	**$21,588,700** This vendor did not provide CR estimates so we estimated costs similar to Vendor 1 in calculating this total.

Estimated costs of computerized science test from Vendor 1 is $2.00 per student. This estimate does not include constructed response items nor items with sophisticated graphics. Vendor 2 did not provide an estimate for the science test.

The following cost estimates are costs that are not included in the above estimations, but would be incurred by the state:

Science Test Development	$1,467,058
State and Federal Staff	$1,389,711
Teacher Review/OEA	$1,477,705
Writing test in 5th grade	$315,162
Writing test in 10th grade	$502,908
TOTAL	**$5,152,544**

(Estimates come from Fiscal Note – 2003 -04 Session. Bill #S2886-0 (R) Comprehensive Assessment Development Moratorium)