Classroom-based Assessment System for Science: A Model

Barbara S. Plake, Buros Team Leader

Chad W. Buckendahl

James C. Impara

Buros Center for Testing

University of Nebraska-Lincoln

Paper commissioned by the Committee on Test Design for K-12 Science Achievement
Center for Education
National Research Council

Acknowledgments

This project was funded in part by a contract from the National Research Council. The project would not have been possible without the support of the NRC's Committee on Test Design for K-12 Science Achievement. In particular, special thanks are offered to Meryl Berthendahl who provide excellent advice for external committee members and to Myron (Mike) Atkins, liaison from the Committee on Test Design for K-12 Science Achievement. The opinions and conclusions stated in this paper are those of the Buros Model Team and do not necessarily represent those of the NRC Committee on Test Design for K-12 Science Achievement.

Buros Center for Testing

Classroom-based Science Assessment System Team

Barbara S. Plake, Buros Center for Testing

Chad W. Buckendahl, Buros Center for Testing

James C. Impara, Buros Center for Testing

Ian MacGregor, Earth Science Consultant

Scott Marion, National Center for the Improvement in Educational Assessment

Carol Shestok, Elementary Science Teacher, Massachusetts

John Mc Kinney, Middle School Science Teacher, Colorado

Teresa Eckhout, Elementary Science Assessment Coordinator, Lincoln, NE

Cindy Gray, Director of Curriculum and Assessment Elkhorn Public Schools

Thomas Neumann, High School Science Teacher, Millard, NE

Arthur Zygielbaum, National Center for Information Technology in Education,

University of Nebraska – Lincoln, Liaison from Technology Team

Technology Team

Arthur Zygielbaum, National Center for Information Technology in Education

Gwen Nugent National Center for Information Technology in Education

Roger Bruning, Professor University of Nebraska

Albert Steckelberg, Professor University of Nebraska

Myron Atkin, Stanford University, Liaison from the NRC Committee on Test Design for

K-12 Science Achievement

Classroom-based Assessment System for Science: A Model

Introduction

States are facing increased challenges as they develop their statewide assessment programs to be in compliance with the No Child Left Behind legislation (NCLB, 2001). Starting in 2002, states were required to implement statewide accountability models based on student achievement in reading and mathematics. By school year 2005-2006 states are required to assess reading and mathematics in grades 3 – 8 and at least once in high school. Statewide science assessments, in at least one grade each in elementary, middle, and high school, must be administered by the 2007-2008 school year. Currently NCLB does not require the science results be included in the accountability system.

With an eye toward providing insights to states on models for statewide science assessments that are consistent with NCLB, the National Research Council's Committee on Test Design for K-12 Science Achievement has asked four teams to propose assessment systems models that may be informative to states as they consider the appropriate approaches for measuring student science achievement. The Buros Center for Testing was asked to build a coherent, instructionally useful, teacher-led assessment program that would meet the various criteria for technical rigor and that would provide adequate reporting at different levels and to different stakeholders.

The model presented by the Buros Center for Testing team uses as its basic foundation the principle that student achievement is the primary focus for any educational program. Therefore, as the model evolved, central in the Team's mind was how this science assessment model could be useful in informing instructional practices and aiding

student achievement. The team recognizes that the chosen approach may not appear psychometrically pristine but it includes creative solutions with high potential to simultaneously improve instructional practice and learning provide information that would be congruent with the requirements of the NCLB legislation. Of course, this presented many challenges and unresolved issues. In some cases, there were no best or recommended strategies for the assessment model. In those instances, the report presents options and implications for those options for (a) meeting the goals of instructional utility and student learning and (b) meeting technical standards.

Overview of the Classroom-based Science Assessment Model

Key features of this model are that the selection of assessments that are used to inform the state on student performance in science are made at the classroom level, often using teacher-developed assessments or assessment strategies. Therefore, teachers are a critical piece in this science assessment model. In order for such a system to be effective and informative at the state level, several fundamental features must be in place. First, teachers must understand the state content standards and incorporate them in their curricular and instructional programs. Second, teachers must be able to gather meaningful and appropriate information about how their students are performing on the state content standards using assessment strategies at the classroom level. Third, procedures need to be identified that will accumulate teachers' reporting of their students' achievement levels on the content standards to support the accountability reporting requirements of the NCLB legislation at the state level. All three of these features

present significant challenges for any assessment system, but meeting them is particularly challenging in a classroom-based state assessment program.

There are important roles for both teachers and the state for the classroom assessment program. To be successful, the assessment program must be an integrated process with common goals for both groups. Teachers need to understand the state content standards and inquiry standards, prepare and use assessments to instruct students, gather individual and longitudinal assessment data that is rigorous and amenable to analyses at the classroom, school, district and state levels. The state must define a coherent and conceptually logical set of standards, train teachers to understand the standard and assessment systems being used, train teachers in assessment methodologies, and assessment preparation and evaluation, and accumulate assessment data to provide schools with helpful information on achievement and information on how to improve the educational system.

Nebraska Plan

These enumerated critical features are in large part in place in the state assessment and accountability system in the State of Nebraska. Nebraska's Student-Centered Teacher-Led Assessment and Reporting System (STARS) was started in 2000 with the goal of using local assessment information to inform state policy and accountability decisions for Nebraska students in Reading and Mathematics. Several features of the STARS program are noteworthy.

First, it was developed based on the strong vision and leadership of Nebraska's Commissioner of Education, Dr. Douglas Christenson. Foremost in the Commissioner's mind was the focus that students are the primary stakeholders of education in the state.

The main goal of the assessment program is to improve student learning through better-informed instructional practices.

Second, performance of students in Nebraska on external achievement tests, in general, is very strong. State National Assessment of Educational Progress (NAEP) results in reading and mathematics routinely place Nebraska students in the top ten states nationally. Although improvement of instructional programs and student learning is the main goal, most educators and policy makers in the state believe that the current instructional programs are, in most cases, working and that students show strong evidence of achievement, particularly in reading and mathematics.

Third, it may be relevant to note that the STARS program was developed and implemented prior to NCLB legislation, so a focus on accountability and direct comparability among districts was not as central as it may be in current and future implementations of state assessment programs.

The Nebraska STARS program has garnered national attention as states work to meet the directives of NCLB. Several articles have been written to summarize the Nebraska STARS. Two Phi Delta Kappan articles (Roschewskii, Isernhagen,, Gallagher, C 2001; Roschewski, 2003) provide an overview of the program and some of its structural features. The summer, 2004 issue of Educational Measurement: Issues and Practice (a publication sponsored by the National Council on Measurement in Education), is devoted to a special issue on the Nebraska STARS. As a broad overview, the Nebraska STARS requires districts in Nebraska to identify their approach to measuring the state's content standards in reading and mathematics. As originally conceived, these assessments would occur in alternating years for reading and mathematics for grades 4, 8,

and 11 (NCLB requirements have necessitated changes in the grades tested and sequencing of reading and mathematics). Districts are required to describe how they intend to measure the state content standards in a District Assessment Plan, which is submitted and reviewed in advance of implementation.

Districts have much flexibility and choice in how the content standards are assessed, but not what content standards they will address (except for some districts that have received approval for their locally developed district content standards under the condition that they are equal to or more rigorous than the state's content standards). In addition, districts have the option of collaborating with other districts to combine their assessment development decisions and resources. Districts can decide to develop a customized assessment program that measures the targeted content standards or they may decide to use a variety of assessment components, each of which provides information on some parts of the content standards. There are two features of the district assessment program, however, that are not optional. As part of state regulations districts must annually administer a standardized nationally norm-referenced achievement test from a predetermined list of tests and they must administer, currently in grades 4, 8, and 11, a statewide writing assessment that is linked to the state's grade-level content standards for writing. Further, Nebraska participates in the statewide NAEP administrations, providing additional evidence of student performance.

Another important feature of the Nebraska STARS is the reliance on an external evaluation of the technical quality of the districts' assessments. Early in the development of the plan, in consultation with the Buros Center for Testing, the state articulated technical quality criteria that were to be considered in evaluating the technical quality of

these assessments. Districts were informed that the assessments used in measuring their students' achievement would be examined to verify that the test results met standards for technical quality. These criteria are:

- Alignment of the test content to the standards
- Evidence that students had the opportunity to learn the content prior to assessment
- Demonstration that the test format and tasks are fair (free of situational bias) and have test content at the appropriate level
- The results show consistency and generalizability
- Performance level cutscores are appropriate

Annually, districts are responsible for compiling an assessment portfolio documenting evidence of how their assessments meet these technical quality criteria. External evaluators rate these quality criteria using rubrics and provide ratings that, when combined, rate the district's assessment program as being either Unacceptable, Needs Improvement, Good, Very Good, and Exemplary. For more information on the technical quality criteria or the evaluation program, see Plake, Buckendahl, & Impara, 2004 and Buckendahl, Plake, and Impara, 2004.

Another critical component that supports the success of Nebraska's STARS is the state's emphasis and support for professional development for teachers and administrators in assessment. Several programs are directed at strengthening assessment literacy and competence of Nebraska educators. One program, sponsored by the Commissioner of Education, is the Assessment Team program that was implemented

statewide under a contract with the Assessment Training Institute. Learning teams were formed and lessons followed under the supervision of Dr. Richard Stiggins. This program was operational from 2000 – 2002. The Department of Educational Psychology's Quantitative and Qualitative Methods sponsors a second program at the University of Nebraska – Lincoln. Intended to lead toward a state approved Assessment Endorsement, cohorts of practicing teachers are recruited into an 18 hour program with the focus on improving teachers' assessment literacy. The program began with coursework in basic assessment principles and practices during the first summer of the program. During the academic year, participating teachers develop local assessment plans and participate in assessment related activities under the direction of the university-based supervisor. The following summer, cohort members continue their learning in assessment through more coursework on advanced topics. To date, three cohorts have participated in the program. A similar program is also being implemented for administrators.

There are important elements of the Nebraska plan that have implications for a classroom-led assessment model. First, local educators are recognized as a central piece in the assessment plan. Decisions about how to approach the assessment requirements are made at the local level. Often, teachers are engaged in the development and administration of the district assessment program. Although not universal, many districts use teacher-developed classroom assessments to measure student performance on the content standards. Second, technical quality indicators of the assessments used in the assessment system are articulated to local educators. Therefore, these quality indicators for assessments are public and serve as the basis for evidence that districts are required to

provide to support the technical quality of their assessments. Third, the need for professional development for Nebraska educators was recognized as a linchpin in the STARS program. Policy makers in Nebraska have provided many resources and programs to improve the assessment competence of teachers and administrators in the state.

Also critical to the success of Nebraska's STARS program is the political and policy environment of the state at the time of its design and implementation. The strong vision and support of the State's Commissioner of Education was a key component. When the state legislature proposed a bill to require a single statewide test for reading and mathematics, Dr. Christenson expressed strong resistance. His passion for student learning and education, and his skill in working with state leaders in the legislature are not to be discounted when considering the transportability of a similar program to another state.

Other key features include the overall high level of achievement of students in Nebraska, coupled with the timing of the design and implementation of Nebraska's STARS prior to federal accountability legislation. Because Nebraska had the opportunity to implement this program prior to NCLB, some of the resistance that might be present in current times was in large part avoided. This is not to say, however, that there was universal enthusiasm by educators and legislators across the state about Nebraska STARS. Resistance appeared early in the implementation by teachers and administrators. Teachers were concerned that their role as a teacher was being challenged by the apparent addition of requirements to their jobs. Administrators did not want to spend teacher time and resources on test development when, in their view, there were "ample good tests

available" for such purposes. Why, some administrators asked, should local districts be saddled with the responsibility to develop tests to measure the same content standards as all of the other districts in the state? Would not one test, developed by experts, to measure these common state adopted content standards be a more efficient approach? What these administrators were not recognizing was the power of having local educators spend the time to ensure that their instruction was aligned with the standards and that their assessments could provide useful and meaningful information about student learning in their classrooms. The spin-off benefits of improved teacher competence in assessment, coupled with their ability to better use assessment results in improving teaching, has proven to be the major positive intended consequence of STARS. This development, however, needs to be further expanded to the educational leaders who are making decisions to support their teachers' efforts.

However, Nebraska STARS has elements that are not optimal for a statewide classroom-led assessment system for NCLB reporting. First, although districts are required to define locally the meaning of student performance categories (e.g., Beginning, Progressing, Proficient, and Advanced), these have not been standardized at the state level. Therefore, one district may set high expectations for students who are performing at the Proficient level (for example) whereas another district may have a more "basic mastery level" perspective on what constitutes "proficient" performance. This causes concern when proportions of students assigned to these performance categories are combined for state level reporting. Second, no evaluation is made of the appropriateness of the content or cognitive demand of these local assessments.

Currently, there is no review of the assessments used at the district level for STARS except to review their technical quality. Therefore, teachers may be using technically sound assessments that match the standards, provide reliable and fair assessments of student learning, but may not meet scrutiny for content soundness. It is tempting to trust that teachers, who are delivering the curriculum and who are entrusted with the education of our children, are able to construct content appropriate tests. However, Nebraska STARS has not evaluated this directly. Further, although meaningful efforts were made to provide professional development and support for districts in improving their assessment literacy and providing assistance, direction, and feedback on their assessments' quality through the portfolio development and review process, many districts were not provided needed support. The geographic diversity of the state, coupled with lack of resources, hindered the ability to provide professional development for many of the districts in the state.

Although there is much to borrow and learn from the Nebraska's STARS to support and inform the development of a model for a statewide classroom-led state assessment program, there are areas where improvements and refinements are needed in order for the program to be suitable for a national model.

Classroom-based Science Assessment System for NCLB Reporting

The model we propose has several important features. The model will pay particular attention to the following features:

- Relationship among standards, curriculum, and instruction

- Role of professional development for teachers and administrators
- Importance of technical quality
- Importance of content quality
- Connections between assessment results and interpretation and use of results
- Reporting strategies, including requirements for NCLB
- Opportunity for accommodations for students with special needs
- Role of technology to support the assessment program
- Capacity of teachers to develop scientifically appropriate assessments that address science inquiry
- Flexibility of the assessment model for gathering multiple indicators of student performance across a variety of assessment strategies

Relationship among standards, curriculum and instruction

When examining student achievement in science as a function of state-adopted content standards, one important element is how well these science content standards have been integrated into the grade level curriculum and instruction. For example, when the National Research Council published the National Science Education Standards, they recognized that these standards did not directly connect with instruction. As part of that publication, the NRC included chapters that provide standards for teaching and for professional development for teachers in science. This represents an important and responsible action by the NRC, recognizing that the NRC's inquiry-focused vision of science would not naturally happen in the classroom without support for teachers in the integration of these content standards into the curriculum and instructional practices.

Although not guaranteed to happen by virtue of having a classroom-led assessment program, the chances for teacher buy-in and commitment to the science content standards, as articulated by the state, are higher when the teachers are responsible for the assessment strategies used to assess student levels of performance on these content standards. Tests that are designed at the state level by assessment specialists may be excellent representations of the content standards for the state. However, if the instructional strategies of the teachers, the persons responsible for delivering the curriculum, are not true to these content standards (particularly when they focus on inquiry in the context of content), then student performance on these state developed tests may not be clearly interpretable. Poor performance on the state developed tests may mean that, despite strong instructional programs tied to the standards, students have not yet achieved adequate levels of performance. Poor performance may also mean that students did not have a reasonable opportunity to learn the content, and therefore the poor performance is more a reflection of the lack of responsive curriculum and instruction than poor levels of student performance. It is not clear, without knowledge of the infusion of the standards into the curriculum, what would be appropriate remedies for poor student performance – better instruction on the wrong content or better aligned instruction?

This is a critically important piece of the classroom-led science assessment system. By virtue of involving teachers in the development of assessment strategies that enable reporting on the science content standards, teachers will be more likely to infuse these science content standards into their instructional practices. There will be a clear and direct link between what happens instructionally and what is assessed. In many

cases, the distinctions between instruction and assessment will become blurred, depending on the assessment strategies used by the teacher.

Just because teachers are more aware of and responsible for assessing science does not mean that they have the needed skills to make the use of the resulting data to adjust their instructional plans and assessment strategies. Therefore, the program of professional development must clearly address the linkages among the science standards, curriculum, instruction, and assessment.

Role of Professional Development for Teachers and Administrators

One of the major challenges in a statewide classroom-based assessment system is the need for working with a diverse set of educators. Although Nebraska, like most states, has a system of regional educational support units (in Nebraska they are called "Educational Service Units", or ESUs), they may not be capable, either from a resource or competency perspective, to provide the level of professional development that would be needed to support a classroom-led science assessment system. In addition to ensuring a tight linkage among standards, curriculum, instruction, and assessment, many teachers will need additional skills in order to develop technically sound and instructionally meaningful assessment tasks. This training need must be accommodated in a professional development program. It should be pointed out that the kinds of assessment tasks that are applicable to classroom-led science assessments are not limited to standard paper and pencil assessments. The types of assessment tasks should be viewed as broad and flexible, to be responsive to the classroom environment and the science standards. For example, an assessment that is administered over several days, that involves students,

individually or in teams, gathering evidence to support a science related inquiry, could be structured to provide meaningful and technically sound results about student performance. It is possible to envision an assessment task that is sufficiently structured and that has clear administration and performance guidance and adequate scoring criteria that can be administered in a classroom. As a matter of fact, this flexibility is enhanced by the use of classroom-based assessments to inform decisions about student learning in science.

Professional development will also be needed to help teachers to make good use of assessment results. By providing teachers and other educators with information on how to link assessment results to decisions about student performance, instructional strategies, and program effectiveness, the utility of the assessment information will be enhanced. Obviously, one of the direct benefits of classroom-based assessments is access to assessment results in a timely manner. This is contrasted with assessment systems that depend on a single, end of year administration, that often report individual student results the following fall term. This does not allow for meaningful instructional decisions or the direct use of the results to assist in student learning for the students on whom the information was gathered. This is an example of where the utility of the assessment approach, classroom-based assessment, was deemed to be a strong benefit over many statewide single assessment programs. While some would argue that the strong standardization and psychometric strengths of a single statewide assessment has appeal, it is clear that the classroom-based model is more advantageous for making direct connections between student performance and learning.

Another component of professional development will be to provide information to teachers and administrators on appropriate communication of assessment results to students, parents, and other stakeholders. Although a clear strength of the classroom-based model is the direct linkage to instruction and learning, it also provides a rich basis for communicating with stakeholders about student learning and school improvement. However, without adequate knowledge of appropriate communications about assessment performance, teachers and administrators may inadvertently provide incorrect or inadequate information. Therefore, another fertile area for professional development for a classroom-led science assessment system is in the communication of assessment results.

Importance of Technical Quality

When a state uses a single assessment to measure the performance of grade level students on state content standards, for example in science, there is only one assessment that needs to be examined for psychometric quality. However, with a classroom-based science assessment system, there are many assessments of many varied types being used to determine how well students are progressing on the state science standards for a particular grade level. These assessments could be paper and pencil, traditional types of assessments. Or, they could include teacher observations of student work, student work products, or extended evaluations based on inquiry, to name a few assessment approaches. These assessments differ in their format, structure, and utility for making decisions about students' levels of performance on the science content standards at that grade level. Therefore, a system is needed to evaluate, and document, the level of technical quality of these classroom-based assessments.

As presented earlier in this report, the Nebraska STARS program utilizes an external evaluation process to examine the technical quality of the assessments used by districts to assess performance of students on the state's content standards in reading and mathematics. Districts are required to assemble District Assessment Portfolios that provide evidence of how the district's assessments meet the six technical quality criteria. These technical quality criteria are:

- Alignment of the test content to the state's content standards
- Opportunity for students to learn the content prior to the administration of the assessment
- Freedom of the assessment from inappropriate content, bias, or sensitive situations
- Appropriateness of the level of content of the assessments
- Score consistency and generalizability
- Appropriateness of performance level cutscores

In advance of submitting their District Assessment Portfolios, districts are given training and preparatory guidelines for how to assemble their portfolios and what kinds of evidence is needed to document that their assessments are meeting these technical quality criteria. Rubrics were developed to be used by the external team in evaluating whether the evidence provided for each criterion is sufficient for a "Met" rating. Across the six criteria, Met and Not Met ratings are combined to provide an overall technical quality rating for the district's assessments. These overall ratings range from Unacceptable to Exemplary. The criteria are not equally weighted in determining the overall rating. The

first two criteria, Alignment of the test content and Opportunity to learn, are essential for receiving at least a "Needs Improvement" rating. Any district whose assessment portfolio receives a "Not Met" on either of these criteria is given an "Unacceptable" overall rating regardless of the ratings on the other four criteria. More information on the rating system and the external evaluation process can be found in the Summer, 2004 Special Issue of Educational Measurement: Issues and Practice (Plake, Buckendahl, & Impara, 2004).

For the classroom-based science assessment system many similarities will be retained with the procedures used to evaluate the technical quality for the assessments in the Nebraska STARS system. Although the assessments will be developed and delivered by the classroom teachers, the district will still be the reporting unit to the state. Therefore districts will be responsible for assembling District Assessment Portfolios that document how classroom science assessments are being used to measure student progress on state science content standards. In some cases, when districts use a unified district assessment model to measure student performance in reading and math, portfolio development will be more complex than what was experienced in the Nebraska STARS portfolio assembly process. This is because schools will need to provide reports about how these classroom-based assessments are being used in the school's classrooms to measure student performance. A district with multiple schools and multiple classrooms at a particular grade level will need to assemble evidence for these six quality criteria for coordination at the district level. Because of this complexity, some variation in the evaluation process for examining technical quality of these classroom-based assessments will be needed.

There will still be a team of external evaluators whose task will be to rate the technical quality of the districts' assessment portfolios using the six quality criteria and overall rating system. However, peer reviewers from districts will augment the external evaluator team across the state. Of course not all districts will be able to have peer evaluators every year, but a rotational system could be implemented such that different districts have the opportunity to participate in the peer review process. Another possibility would be to engage staff from the regional ESUs from across the state to serve as peer reviewers. Peer reviewers would have responsibilities at the district level to aid in the assembly of districts' assessment portfolios and in rating portfolios from other districts. An issue that would need attention, however, would be if there were any rewards or sanctions implemented for districts, especially in a competitive sense, based on the districts' ratings on their assessment portfolios. No rewards or sanctions, competitive or otherwise, are currently in place in Nebraska (aside from publicly reporting of performance). However, if this situation were to change, a reconsideration of the role of peer reviewers in rating other districts' assessment portfolios would be needed.

The results of the technical quality ratings are made public. In the Nebraska STARS reporting of student performance at the district level, the district's rating on technical quality appears juxtaposed to the performance rating. Therefore, a district with high levels of performance, but based on poor quality assessments may find that their student performance results discounted or at least questioned. On the other hand, a district with high ratings on technical quality presents student performance that is more trustworthy and results that are more likely to be interpreted meaningfully.

Serious challenges need to be addressed due to the magnitude of the task in evaluating and monitoring the technical quality of classroom-based science assessments used to measure student performance on state content standards. It might be tempting to reject this classroom-based science statewide assessment system based simply on the magnitude of the task of compiling and examining the technical quality of the assessments used at the classroom level. This is a place where the goal of student learning and instructional improvement was elevated in weighing the positives and negatives of such a classroom-based assessment system. Certainly the need for quality assessments should not be diminished, as the veracity of the assessment results rests on the integrity of the assessments used to obtain those results. However, in balance between the management of a classroom-based system and the need to have a more streamlined assessment program (such as would be the case with a single statewide science assessment at a particular grade level), the positives from the classroom-based system appear to outweigh the logistical and management concerns for implementing such a system.

Importance of Content Quality

The technical quality of the classroom-based assessment tasks must be examined with consideration of critical content validity requirements. It is tempting to trust that teachers, by virtue of their training and their central role in curriculum and instructional planning and delivery, would be in a position to develop content appropriate science assessments for their grade level. While this is probably a correct assumption for many teachers operating within their major areas of training, it may not apply in cases where

the teacher is teaching outside his or her major area of training. Further, especially if the NRC National Science Standards is the basis for a state content model, some of the emphases in these standards, such as the role of inquiry, may not have been a dominant component in the science teachers' preservice training. Part of this concern has already been addressed in the section on the integration of the content standards with curriculum, instruction, and assessment.

Therefore, it is necessary that some content review be provided of the classroom-based science assessments that are used for measuring student performance on the science content standards. This might be folded into the District Assessment Portfolios, with the external and peer reviewers' membership being expanded to involve state science coordinators, science teachers, and science experts.

Connections among assessment results and interpretation and use of results

A strong positive benefit of the classroom-based science assessment model is the direct connection among curriculum, instruction, and student learning. This has been highlighted earlier in the discussion of the need to enhance teachers' infusion of the science content standards into their instructional and assessment strategies. However, this synergy is not fulfilled unless the results of the assessments are used to make meaningful interpretations of student performance in order to improve student learning and instructional strategies. It is not always obvious to teachers, students, administrators, and parents how to best use assessment results to make appropriate decisions about student performance and the effectiveness of instructional programs. As discussed earlier in the report, professional development activities will need to be designed to provide

additional support in the use of assessment results to make meaningful decisions. Because the classroom-based science assessment system involves classroom assessments administered as a result of an instructional program, it would be very disappointing if teachers were not skilled in how to capitalize on this source of information to make improvement in student learning through instructional practices. A major benefit of the classroom-based aspect of this model is that tests are being developed by students' teachers as part of students' regular classroom experiences. This is in contrast to other assessment models where the test is administered at the end of the school year, probably at a time that is instructionally distant to the actual student learning of the content. Under that plan, test results often are not available to the students' classroom teacher until the next year, when the student has been promoted to another grade or moved to another school. It is obvious that the opportunity for meaningful use of the assessment results is richer for the classroom-based assessment system. However, the appropriate use of the assessment results will not be realized if teachers do not have the needed skills in interpretation and use of these classroom assessment results. Therefore, it is important that information about ways to interpret and use classroom assessment results be a major component of the professional development effort for the classroom-based science assessment program.

Reporting strategies, including requirements for NCLB

There are major challenges for a classroom-based science assessment system to meet the reporting requirements for the No Child Left Behind Legislation. States must be able to report, for the full group of test takers, the percentage who are performing at or

above the "proficient" achievement level. For a centralized assessment system, like those that use a single statewide test, the task is to identify what is the minimum performance level on the test that is consistent with the knowledge, skills, and achievements for students functioning at or above the proficient level. This entails the creation of a test that is tied to the science content standards, articulation of the knowledge, skills, and achievements of a proficient level student, and then undertaking a process (called a standard setting study) to determine the score on the test that best differentiates non-proficient test takers from proficient test takers. Similar procedures would be used for determining the cutscores that identify other levels of performance, such as Basic and Advanced.

There are several key features of a standard setting process with a single test that need to be highlighted when considering using such a process for a classroom-based science assessment system. First, with the classroom-based assessment system, there isn't a single test that is used to measure student performance on the science content standards at a grade level. Many tests are used, at different times, and of different types. All of these tests have the goal of measuring student performance on the state content standards, but there is no specific commonality across these assessment tasks other than the state science content standards. Because there is no single test given at one time, it isn't feasible to have a single standard setting effort to identify the specific score values that would represent student level performance for Basic, Proficient, or Advanced. Obviously, this presents a serious challenge for a classroom-based science assessment system.

Central to the plan for reporting student performance to meet NCLB requirements is the necessity for a statewide endorsed statement of the knowledge, skills, and achievements that define student performance for Basic, Proficient, and Advanced in science for each grade levels. Teachers will be charged with the task to develop classroom-based science assessments that not only target the state content standards at their grade level, but that also provide sufficient breadth and depth of coverage to be able to differentiate Basic, Proficient, and Advanced levels of performance in their students. This may be especially challenging for some teachers who have historically relied on external, rather than classroom assessments, to make these kinds of distinctions. Many teachers prepare tests and assessment tasks that they believe should be answered correctly by each student in their classes. They are looking for evidence that the students have learned the concepts presented through direct instruction, not necessarily to examine whether students can make the kinds of extensions of the material that often define advanced levels of performance. We see this as a major focus for professional development for teachers.

In order to meet the NCLB reporting requirements, teachers will need to track whether each of their student's level of performance is consistent with Basic, Proficient, or Advanced levels. It is essential that these teachers understand the importance of setting appropriate performance level criteria for their classroom-based assessments. Some teachers may elect to classify students into performance categories based on their end of unit tests. If this is the case, then teachers will need to be sure that these end-of-unit tests have sufficient match to the content standards and breadth of complexity in order to appropriately classify students into the multiple performance categories. In

addition, sufficient depth of coverage will need to be built into the assessment so there are sufficient opportunities to demonstrate performance at a given level. Further, teachers must make certain that the cutpoints they employ for making these determinations are sensitive to the difficulty of the tasks on the assessment, and not set arbitrarily using pre-specified percentage correct grading standards. For example, many districts have grading policies for A, B, C, D, and F grades. Many times these grades are based on the percent correct on the test. It would be inappropriate, for example, to decide that Advanced performance would always be based on achieving 95% correct on the classroom-based test. It is feasible that a teacher might develop a test to assess the relevant science content for that grade level, which would require a very basic understanding of the content, resulting in achievement of the 95% criterion; likewise a test could be constructed over the same content where even students with Advanced levels of understanding of the content would be hard pressed to earn this percent correct. In order to set appropriate performance level cutscores, a decision must be made based on the match between the stated performance level descriptors for the various performance levels and the cognitive and content demands on the tasks that comprise the assessment.

The situation is made even more challenging when a teacher decides to weigh an accumulation of student work to decide whether the student is performing at the Basic, Proficient, or Advanced performance levels. A teacher may decide to use a collection of student work, accumulated across a science unit, to make student performance level classifications. This requires the teacher to have a systematic approach to gathering, evaluating, and classifying student work based on the collection of student performance. Student portfolios may be useful as a strategy for teachers to collect, organize, and

evaluate an accumulation of student work across an instructional unit. Rubrics would need to be developed that specify what evidence the student needs to provide, and at what levels of complexity, in order to warrant classification into the different performance categories. This evidence would become part of the teacher's, school's, and district's assessment portfolios that would be examined for both technical and content quality.

Another critical feature of a classroom-based science assessment system is the record keeping at the teacher, school, district, and state levels. In order to provide the needed information for NCLB reporting, states will need to provide districts, schools, and teachers with a standardized data structure to use in reporting student performance. The resulting database must support all of the disaggregations that are dictated by the NCLB legislation.

Simply put, teachers will have the responsibility to classify their students, based on their students' performance on classroom assessments, into multiple performance categories using a procedure that is consistent with the state's performance level descriptors and the rigors and complexities of their classroom tests. It is left up to the teacher to decide a) whether to base these decisions on a single unit test or on multiple assessment tasks; and b) whether to allow students to have multiple opportunities to show their achievement of these pre-defined performance levels. What is not discretionary is what content is to be covered (this is defined by the content standards) or what level of performance is required for classification into the performance categories (this is defined by the performance level descriptors). Further, teachers must have a mechanism for reporting to the more centralized unit (school) the classifications of their students into the performance categories. Further, teachers will be required to document the process they

used to make these decisions; this documentation will become part of the teachers' assessment portfolio.

Once teachers have reported student performance, students who are classified into the same performance level will be assumed to be able to perform similarly on science content tasks that tap performance at that level. Therefore, the accumulation across the state of the numbers of students assigned by their teachers into the performance levels will provide a mechanism for meeting the reporting requirements for the NCLB legislation. It is the uniform meaning of the performance level descriptors that serves as the common link for reports from teachers across the state. It does not matter what mechanisms were used to measure proficient levels of performance; they are allowed to vary in this classroom-based assessment model. The common component is that, however measured, all of these assessment approaches are able to report results that allow for the classification of students into the performance categories based on the performance level descriptors.

Clearly, this is a weak link for making aggregated classification decisions. Stronger approaches using well-known equating procedures are not feasible due to the complex and diffuse nature of the classroom-based assessment system. For example, strict assumptions would need to be met about the distributions of test scores and the use of common test questions for equating. There is no way that such a controlled approach could be applied for a classroom-based assessment system. Instead, the system relies on a far weaker set of assumptions and less certainty in the common meaning of student classifications. Again, this was a place where the choice between approaches was weighed. And again, this was a place where it is our decision that the gain in equality of

meaning for the classification decisions was viewed as not equal to the gain in instructional utility of the system that is based on classroom assessment, tailored to the instructional experiences of the students and administered in instructionally meaningful time.

Although it is not expected at this time that NCLB requirements for science reporting will include trends across years, this requirement could be met through the same conceptual process of linking performance across time by using the common meaning of the performance level descriptors. Although the assessments are permitted to vary across classroom, the meanings of the performance levels descriptors are not. Therefore, trends across time will be linked through the common meaning of the performance level descriptors. The assessments will change (but there are already multiple assessment strategies in place for a single year's reporting) from time to time, but across years the meaning of the performance level descriptors will remain constant. If the performance level descriptors were changed, no equating or linking strategy would be helpful in maintaining trend information. To compute trend data, the numbers of students who are classified as performing at the "Proficient" level or above would be tallied and then converted to a percentage. Every year when the data are reported by the districts, the percentage of students who are performing at least at the Proficient level will be calculated. Because the veracity of these results are dependent on using technically sound instruments, the ratings of the districts' assessment portfolio should be considered when aggregating the results.

For example, if a district's assessment portfolio receives at least a "Good" rating, the numbers reported by the district could be directly aggregated into the state's totals for

students who are performing at Proficient or above. If, however, the district's assessment portfolio has a rating lower than "Good", that means that the assessment results are not trustworthy and potentially invalid. Some decisions would need to be made about how to treat assessment results from these districts and would need to reflect the expectations for assessment quality that are likely to result in trustworthy performance estimates. One such policy decision is to discount the numbers of students who are reported in the performance level categories (see Buckendahl, Impara, and Plake, 2002 for one such discounting model). Once these baseline totals have been established, change in this percentage value would indicate whether adequate yearly progress is being made in science.

Reports will need to be constructed to inform various stakeholders of the performance of students in the state on the state content standards. Depending on the stakeholder, different information should be provided. For example, at the classroom level, students, parents, and teachers are the primary stakeholders for these results. The reports should focus on progress the student is making in learning the science content standards and where weaknesses are noted. This means that the assessments will need to be detailed enough to allow for meaningful diagnostic reporting. Also at the classroom level, teachers will want to have information on the effectiveness of the instructional units that were planned and delivered to cover the science content standards. Teachers will need professional development experiences to enhance their understanding and ability to provide diagnostic information to students and parents and to gain relevant information about instructional effectiveness. Some stakeholders will want to use the results to compare how well students are performing across districts within the state.

This can be accomplished by comparing the percentage of students from the districts who are classified into the four performance levels; particular attention may likely be placed on the percentages of students who are not yet performing at the state defined Proficient performance category and the percentage of students who are in the Advanced performance category. This comparability is built into the system through the establishment of state defined performance level descriptors that are applied consistently across the districts in the state.

For reports from the state that summarize student performance by districts and schools, information should be reported about the ratings on technical and content quality as that information is useful for stakeholders when interpreting the performance ratings reported by the districts. In order for district results to be meaningful, a minimum level of technical quality must be maintained. Districts' technical quality ratings of "Unacceptable" or "Needs Improvement" do not support valid interpretations of student performance. The state will need to take corrective action in order to ensure that all district assessments are rated as at least "Good." Otherwise, their quality will not be sufficient for reporting in aggregate for NCLB.

Opportunity for accommodations for students with special needs

Most states with a centralized statewide assessment system have decided to use the students' Individualized Education Plan (IEP) to form the basis for any decisions about accommodations that might be permitted during the statewide assessment administration. Because these centralized statewide assessments are often administered within a limited time period, often outside of the student's regular classroom setting,

students identified to receive accommodations may be placed in a different location with different test administrators who are not familiar with the student or his or her particular needs (beyond that specified for accommodated administrations). A classroom-based assessment system has the promise of providing a more familiar and routine assessment experience for students who need adapted or accommodated assessments. These assessments would be administered in a "business as usual" context; same classroom, same teacher, and same accommodations that are specified in the student's IEP. This has the potential of providing a more comfortable circumstance for measuring these students on the state science content standards. Therefore this is seen as a strength for a classroom-based science assessment system as it is more "student friendly" not just for students with special needs, but for all students. This could be especially beneficial for students who have not been identified as needing an accommodated administration but who feel particularly anxious in formal testing settings.

Equally important, this assessment program could be well tailored to gifted and talented students as well. By requiring that the assessments permit assessment of higher levels of learning for advanced levels of performance, students who have gone beyond the proficient levels of achievement can be identified and their performance monitored. In some state assessment programs, where the emphasis in on measuring proficient levels of performance (with the motivations to show Annual Yearly Progress for students being classified as proficient or above) some worry that the stronger students who are already performing at the higher proficiency levels will be "lost in the shuffle". A classroom-based assessment system holds promise to keep the achievement of these gifted students in the forefront of the teachers and the state.

Role of technology to support the assessment program

Because there are so many possibilities for how technology could be used to support a classroom-based science assessment program, a special advisory committee was formed to consider these possibilities. The Classroom-based Science Assessment Technology Group comprised Arthur Zygielbaum, liaison to the Buros Team, Roger Bruning, Gwen Nugent, and Al Steckelberg. This group meet with Barbara Plake, Buros Team Leader, on two occasions to examine possible directions for the use of technology in a classroom-based science assessment system. Five major areas were identified:

- Use of technology for delivering professional development
- Use of technology in the preparation, rating, and storing of district assessment portfolios
- Use of technology in the creation and delivery of science assessments
- Use of technology for reporting of assessment results
- Use of technology as a conduit for communication

Use of technology for delivering professional development. As noted above, professional development is seen as a critical piece for the success of a classroom-based science assessment system. Part of the challenge in conducting these professional development activities is reaching all of the participants across a wide geographic region. The use of technology to deliver distance-learning modules that address the professional development needs was seen as an efficient and effective approach. As mentioned before, there are several areas where professional development has been highlighted as a

needed element of this model for providing critically important skills and knowledge to stakeholders in a classroom-based science assessment program. To repeat a few, it was mentioned that teachers would likely benefit from professional development that would help them with their integration of the content standards into their curriculum, instruction, and assessments. For some teachers, additional training in science content might be warranted, especially for areas in the content standards that were not covered in as much depth in the preservice teacher education programs (such as the emphasis on inquiry that is present in the National Science Standards published by the National Research Council). Further, it was stated that focused professional development might be helpful for teachers to make the needed connections between assessment results and use of these results in improving student learning and for making instructional decisions.

Teachers would likely benefit from professional development that is aimed at increasing their assessment development skills. Many teachers did not have an assessment course as part of their preservice education. Many teachers are not up-to-date in some of the more recent assessment techniques and strategies that might be particularly useful in a classroom-based assessment system, such as performance assessments or the use of portfolios in assessment. In addition, teachers will be required to gather evidence about how their assessments meet the six technical quality criteria; so specialized professional development on the meaning of these quality criteria will be needed. Moreover, follow-up professional development for these teachers in ways they can document the technical quality of their assessments would probably be helpful to demystify these technical quality criteria. Many teachers will want to consider using tests that were created by others, such as the unit tests that appear in classroom textbooks.

Teachers may not be informed users of these tests and lack the needed skills to be critical of the quality and utility of these pre-prepared tests. A professional development module that addresses how to evaluate the quality of such assessments might be especially beneficial.

The need for assessment literacy does not end at the classroom door. Assessment results are communicated to both students and parents, who are likely ill prepared to make appropriate interpretations of the test results. Parents and students may be interested in distance learning modules tailored to their interests in how to interpret test results, how to connect test results with classroom instruction and learning, and in ways they may help their students maximize their performance on tests (helpful hints for test taking strategies).

Administrators would benefit from professional development that is aimed at communicating test results to parents, board members, and journalists. They would also find helpful, we expect, a professional development experience that was directed at understanding disaggregated test results, especially with regard to the NCLB reporting requirements. Based on the hectic schedules that most school administrators face, having these professional development experiences available to them at their convenience through a web-based delivery system may be attractive. A user-friendly interface would also facilitate the necessary record keeping.

Policy makers should not be left out when information about assessment is being prepared for delivery to assessment stakeholders. Many times policy makers view tests as a quick and easy solution to educational problems. Most of these policy makers have good intentions but limited knowledge about appropriate and inappropriate uses of tests.

With a technology based system, special modules could be developed for policy makers to help them become better informed on appropriate test use and on how to interpret and communicate test results.

Therefore, in the area of assessment literacy training, many opportunities are envisioned for the use of technology to deliver focused assessment-based information to a variety of audiences that fulfils two important objectives: (a) it is available to a wide audience of users and (b) it is available at their convenience. This will increase the reach of professional development access to teachers and educators across the state. Further, because educators and policy makers are often time-restricted, it makes the professional development program meet their schedules, not requiring them to take time out of their schedules to attend professional development workshops that may require additional travel and other inconveniences.

Use of technology in the preparation, rating, and storing of district assessment portfolios. Another area where technology holds promise for use in a classroom-based science assessment system is with the District Assessment Portfolios. As stated previously, these portfolios will be rated on technical and content quality by external and peer evaluators. The process of assembling a District Assessment Portfolio can be a daunting task for districts. There is little structure to the system used by the Nebraska STARS program, which makes the assembly cumbersome and often confusing task for district personnel. District Assessment Portfolio reporting that is supported by technology could provide both this structure and cut down on the accumulation of paper that is currently needed to document evidence for the six technical quality criteria.

The Nebraska STARS program currently has in place an electronic rating form that is used by the external evaluators when they rate the District Assessment Portfolios. This electronic system has radio buttons for each of the criteria that identify common problem areas for the specific quality criteria. In addition to recording their rating for each of the technical quality criteria, evaluators can select from these pre-determined problem areas the ones that are appropriate for their evaluation of the quality criteria. Text boxes are also available that allow the evaluators to provide direct, specialized feedback to the districts about the rating on each of the criteria. Any time a rating other than "Met" is given, the evaluator is required to provide individualized constructive feedback on what steps the district needs to take to rectify the problems. Further, the system is preprogrammed to make the overall rating for the district based on the rubric that establishes the district performance of these six quality criteria for ratings of Unacceptable, Needs Improvement, Good, Very Good, and Exemplary.

Another benefit for creating electronic portfolios is in the logistics of making them available to the external reviewers and in their long-term storage. Currently with the Nebraska STARS evaluation effort, a team of 15 – 20 Ph.D. level evaluators is brought to Lincoln, Nebraska for a 3-day training session every summer. Then later in the summer, multiple boxes of District Assessment Portfolios are sent via overnight delivery to the evaluators for their review and rating. When their ratings are completed these District Portfolios are then overnighted back to the Buros Center for Testing for their use in the appeals process. Portfolios are stored at the Buros Center for Testing until the appeal process is completed and then transferred to the Nebraska Department of Education for return to the districts. The awkward and costly process of shipping to and

from the evaluators and storage in Buros's and the Nebraska Department of Education's offices would be avoided if the District Assessment Portfolios were electronically prepared. It would also provide for easier recovery of materials if a portfolio is lost, damaged, or destroyed.

Another use of technology to support the District Assessment Portfolios would be the creation of electronically accessible models that show acceptable evidence for the quality criteria. Other districts that may be struggling to prepare their portfolio may find such models very helpful.

Use of technology in the creation and delivery of science assessments.

Technology can also be a useful tool in the creation and delivery of science assessments. Although it is encouraged that teachers use assessments they have personally developed in this classroom-based science assessment system, it is expected that teachers will want to collaborate on the development of assessment tasks. Technology could be provided that would assist in this collaborative effort. Electronic clearinghouses and chat rooms could be set up to assist teachers in finding support, ideas, and assessment tasks that might be helpful to them in their classroom assessment strategies.

In addition to possibly providing a support mechanism for teachers to collaborate on assessment development in science, technology could serve as a medium for the delivery of science-based simulations. This might have particular application in measuring inquiry, as the assessment could be programmed to tailor the process based on the student's previous responses, presenting data and outcomes to the student for his or her consideration in a process of scientific inquiry.

A web site could also help by providing operational definitions and descriptions of different levels of performance. It could show assessments that are appropriate, for example, to measure proficient performance and then show examples of proficient performance.

Furthermore, technology could perform an important, but somewhat mundane role in recording and monitoring student performance on the science content standards. This could be especially useful for a teacher whose assessment plan involves the accumulation of evidence across a number of assessment tasks before making a classification decision about the performance level category for his or her students.

Use of technology in the reporting of assessment results. The reporting requirements for NCLB will be supported by an electronic system for recording and summarizing assessment results. In addition to meeting the NCLB reporting requirements involving trend data and disaggregated performance information for special populations, an electronic system could assist in providing stakeholders with results, as needed, for students, schools, districts, and states. Of course all of the important confidentiality and privacy features must be in place in order for the public to feel comfortable with such access and reporting of student performance.

Use of technology as a conduit for communication. Communication is important for any assessment program, whether it is a decentralized approach such as the Nebraska STARS or a centralized system based on a single statewide test. However, because there are so many people involved in the development, delivery, scoring, reporting, and use of

assessments in a decentralized system, effective communication plays an even greater role in a successful system. By having a single, web-based communication system available, communications can be more easily maintained. Such a communication system would provide new information about the assessment system as new federal or state regulations come on board, information about requirements and timelines for preparing and submitting assessment related information, and timelines and schedules for professional development activities. This communication link is viewed as essential for an effective classroom-based statewide assessment program.

Capacity of teachers to develop scientifically appropriate assessments that address science inquiry

Because classroom teachers are the source of assessment tools for measuring their students' performance on the science content standards, one critical issue to be addressed is whether teachers can, in practice, actually develop technically sound and content appropriate science assessments. In order to address this question, a team of practicing science teachers was asked to join the Buros Center for Testing's Model Team. Through negotiations with the NRC Committee on Test Design for K-12 Science Achievement, the Buros Team was given the content area from the NRC National Science Standards of Earth and Space Science. When inviting teachers to join, the Team focus was placed on elementary, middle, and high school teachers in these content areas. A total of 4 science teachers became part of the Buros Team: Carol Shestok, elementary science teacher from Massachusetts, John McKinney, middle school science teacher from Colorado, Teresa Eckhout, elementary science assessment coordinator from Nebraska, and Tom Neumann,

high school science teacher from Nebraska. These teachers (and others) were invited to a meeting of the Buros Team held on February 26, 2004 in Lincoln, NE. Each teacher was asked to bring to the meeting teacher-developed earth or space science assessments as well as assessments that showed the assessment of inquiry in the context of earth or space science. In addition to the assessments supplied by these teacher members of the Buros Team, additional assessments were obtained from Omaha Public Schools and Lincoln Public Schools. Also invited to join the Buros Team were Ian MacGregor, earth science specialist, Cindy Gray, a middle school director of curriculum and assessment, and Art Zygielbaum, a space science specialist (and also serving as the liaison to the Buros Team from the Technology Group, see above). As part of the meeting on February 26, these example teacher-developed assessments were shared and discussed with an eye toward finding illustrative assessments that might demonstrate the capacity of teachers to develop high quality earth or space assessments for use in a classroom-based assessment system. Several assessments were identified as promising for use as illustrations for teacher developed assessments. The goal was to find both content specific and inquiry in the context of content classroom assessments that teachers developed for use with elementary, middle, and high school students. Ian MacGregor provided additional follow-up evaluations of the subset of assessments that were identified from this meeting as promising. The purpose of this activity was to gauge how well, even without the intensive professional development program in assessment literacy, classroom teachers who produce technically and psychometrically sound assessments. Overall, the committee was encouraged by the assessments they reviewed but recognized that this

provided only a very limited evaluation of the capability of classroom teachers to produce acceptable assessments for use in a classroom-based science assessment system.

Flexibility of the assessment model for gathering multiple indicators of student performance across a variety of assessment strategies.

By allowing classroom teachers to be the focal point for the development, delivery, scoring, and use of assessments for measuring student performance on the state's science content standards, many assessment benefits are realized. Classrooms allow for much more flexibility in the kinds of assessment tasks that can be feasibility incorporated in the assessment system. Although this point has been alluded to in other sections of this report, it is emphasized here as a major advantage of the classroom-based science assessment system.

Many educators have a very limited view of what constitutes an assessment task or event. Many teachers view an assessment as a paper and pencil, often multiple-choice test, because of their limited exposure to assessment theory and practice in their preservice programs. Few educators understand that the framework that underlies an assessment event is the important feature of the task, not the specific format or structure of the assessment itself. The underlying framework dictates that an assessment event must be a systematic observation of behavior with a mechanism for scoring. Therefore, an assessment event could be a teacher-based observation of student science performances, a series of classroom tasks that overall measure students' skills at inquiry, and a portfolio of examples of student work that will be scored using a rubric. Teachers could monitor students' cognitive processes as they proceed through an inquiry task,

evaluating when, or if, students make adjustments in the expectations of outcomes based on preliminary evidence from the inquiry process. These are examples of more flexible assessment events that might be more feasible to include in a classroom-based assessment system. These types of assessments are more challenging to accomplish because they are time consuming and require clear documentation of systematic procedures. They are virtually impossible to implement in a large-scale assessment program because of concerns about time, resources, and control. However, in the classroom, these innovative assessment strategies become more feasible, although still challenging.

Strengths and Weaknesses of a Classroom-based Science Assessment System

This document has presented the basic features of a classroom-based science assessment system that could be used for a statewide assessment system that meets the federal requirement of the No Child Left Behind legislation. A key feature of this assessment system is that is has a "bottom-up" structure. Rather than being an assessment system that is imposed by the state, local classroom teachers serve as the main implementers and assessors of the science standards. They do this through their curriculum planning, instructional design, content deliver, assessment development and scoring, and their interpretation, communication, and implementation of assessment results. The overarching goal of this assessment system is to help inform the educational process so that students have better opportunities to learn. The focus of this assessment system is on student learning; it is the kids and their learning that matters.

A classroom-based assessment system has many important strengths and several critical challenges. This section of the report highlights these strengths and documents the recognized weaknesses.

Strengths of a classroom-based science assessment system

As stated throughout this report, a classroom-based assessment system has at its center classroom instruction and student achievement. Teachers are better able to incorporate the science standards into their instruction and students are given direct and frequent opportunities to both demonstrate their learning and receive extra help in areas of weakness. This process of instruction, learning, assessment, and redirection is typical in a classroom environment. This cycle directly supports learning through the close connections between assessment and feedback to students regarding their learning needs. This science assessment system, therefore, is not an add-on to the regular classroom process.

Through their direct involvement with the science content standards and as a result of their professional development experiences that focus on aiding teachers in understanding and incorporating the science content standards into their instructional practices, teachers will be better prepared to provide rich and appropriate science educational experiences to their students. Teachers are empowered in this system; they are the creators not the recipients of the assessment strategies to measure students' performance on the science content standards. Therefore, the assessment results will have more use and meaning for them because of the relevance of the assessments and the immediacy of the results. Further, it is likely teachers will be more committed to the

science assessment process because they have been given the opportunity to be a part of that process. More teacher buy-in, therefore, is expected.

Another important positive benefit from a classroom-based science assessment system is the growth expected for teachers and other educators in assessment literacy. Teachers will be able to construct better assessment tools, and therefore obtain better information about how their students are learning in their classrooms. Administrators will be better prepared to explain assessment results to community leaders and school board members. Policy makers will have better knowledge of appropriate uses of tests.

Another benefit is related to cost. Expenses related to centralized test production, publication, distribution, scoring, and reporting are diminished or avoided. Costs of developing assessments are widely distributed and in most cases are included in the regular costs of instruction.

Weaknesses

These positive outcomes do not come without substantial challenges. By placing the assessment responsibility in the hands of classroom teachers, additional burdens are realized. Teachers will need to become strong assessment developers. Further, some teachers may need better training in science content before they can develop adequate assessments to measure content knowledge. Some may argue that assessment development should be placed in the hands of the "testing experts," not classroom teachers. Test companies make it their business to develop high quality assessments for use in large scale, statewide assessment programs. How can it be reasonable, some may ask, to expect classroom teachers to produce tests of the same quality? When considering

this important question, we focused again on the main purpose of the assessment program, which is to improve student learning. How do these externally developed, state delivered tests satisfy the goal of improved student learning? Often they use only a limited scope of assessment strategies that do not fully tap the complexities represented in the content standards. They are limited in their ability to do so by financial, logistical, and psychometric constraints. Further, statewide assessments are often administered in a restricted time window that does not correspond with instruction and results are slow to return to the schools, teachers, and students. It is not uncommon for student results to be made available in the fall semester of the next academic year, far too late to make meaningful instructional interventions. So, yes, we agree that professional test developers may have better skills, at least at the start, than do classroom teachers. However, on balance, we firmly believe that classroom teachers should be the ones monitoring student progress on the state's science content standards because they are in the best position to receive and use the students' performance information.

In a similar vein, another weakness in the classroom-based science assessment system is a lack of control over the technical and content quality of these assessments. With a single state test, evaluation of the test's technical quality can and is a routine part of the assessment program. Even though the classroom-based science assessment model proposed here has quality control components imposed through the evaluation of the technical and content quality of the teachers' assessments, it is unlikely that quality assessment tools will be applied uniformly throughout the system by all teachers on all assessments used to measure student performance on the state's science content standards. There are two reasons for reassurance on this point, however. First, no high

stakes decisions about students are intended from the results of the classroom-based science assessments. If, on the other hand, student results were to be used for graduation eligibility or other systems of awards and sanctions, a decentralized classroom-based assessment system would probably not stand up to intense psychometric or legal scrutiny. Second, there are opportunities to corroborate the results from the classroom-based science assessment system with other assessment results. One way to verify the results that are being reported through a classroom-based system is to compare student performance to other assessments administered in the state. For example, the state may participate in the State NAEP program. There should be some congruence between students' reported performance on the classroom-based assessments and how they do on NAEP. Of course there are important differences between the population, motivation, content specifications, and performance level descriptors between NAEP and the classroom-based assessments. However, if the classroom-based approach, for example, showed that a majority of students across the state were performing at the Advanced level (as defined by the state's performance level descriptors) whereas very few were performing at that level on NAEP, some follow up analyses would be in order. A first place to start would be to compare the state and NAEP's performance level descriptors for "Advanced." If they are consistent, then other reasons for this disparity should be examined. Other external assessments could also provide collateral information to help confirm or raise questions about the states' reporting of student performance from the classroom-based science assessment system.

Another area of weakness for the classroom-based science assessment system is the weak evidence for consistency in meaning for classification of students into the

performance categories, such as Basic, Proficient, and Advanced. Although there will be clear statements of the skills, competencies, and achievements that are expected by students who are classified into these performance categories, when used by teachers across the state variances in interpretations can be expected. In order to form meaningful aggregations of performance into these categories for statewide reporting, the lack of consistency in the interpretation of these content category descriptors can be a serious problem. One way to diminish the concern for differences in interpretations for meanings of these performance level descriptors is to have many and clear concrete examples of student work that characterizes these performance categories, These could be mounted on a website so that teachers have easy access to reminders of the level of student work that characterizes the performance categories. Professional development programs should feature the meaning and interpretation of the performance categories with ample practice for teachers to apply their training to demonstrate their ability to make appropriate classifications of student work into these performance categories.

Even with the high level of professional development and excellent training tools for assisting teachers in their internalizing the meaning of these performance level descriptors, it is likely that differences in application will still be a problem. To the extent that students are inaccurately classified into performance categories, reports by the state to meet NCLB requirements will be problematic. Again, there are external assessments that can help lend credence, or instill doubt, in the state's reported results.

Further, this may be an area where decisions of what is most important for the assessment program again are weighed in making an operational decision. The kinds of tests that lend themselves to higher assurance of equated student level performance across

instruments would require far greater control over the assessment process than is feasible with a decentralized, classroom-assessment model. If the decision that strict equal category meaning is of higher value than the student learning and instructional utility achieved in the classroom-assessment model, then the right decision would be to reject the classroom-based assessment system. However, we strongly believe that the benefits of the classroom-based assessment system are far more useful to student learning than would be strict assurance of category meaning. There would be instances where the balance may tip in favor of strict comparability, especially for tests with high stakes consequences for students, schools, districts, and states.

Another very important weakness of the classroom-based system is its cost, both in financial and human resources. This cost, however, is likely lower than the financial outlay to conduct a comprehensive centralized custom statewide assessment program. Teachers will need to participate in an aggressive, on-going, professional development program aimed at enhancing their skills in a number of critically important areas. These include programs designed to improve their understanding of the science content standards and their integration into curriculum plans and instructional strategies. Teachers will also need strong professional development programs to enhance their assessment literacy, including their ability to make good instructional decisions based on assessment results. Teachers will also benefit from professional development programs that increase their skills in communicating assessment results to students and parents. Other professional development efforts to support the classroom-based assessment system are directed to administrators and policy makers.

Although the price tag for these professional development activities is admittedly large, the benefits are huge and enduring. Professional development of this kind can be especially meaningful and useful to teachers because it is directly related to improving their capabilities as teachers. Also, the costs are lower in future years as the technology start-up costs will not continue at the initial rate. By better understanding the science content standards, they will be better science teachers. By knowing more about assessment and assessment practices they will be able to use better assessment strategies which will result in their having better information about the performance levels of their students. Through a better understanding of how to use assessment results, teachers will be able to apply improved instructional strategies to assist in their students' learning. By having better communication skills, these teachers will be able to explain in coherent ways to their students and parents about student progress in their classrooms. A probable benefit is better support for students and teachers by these parents. The outcome of this teacher focused professional development is directly tied to the overall goal of the classroom-based assessment system: better student learning.

The classroom-based science assessment system has other costly elements, including the external evaluation of the technical and content quality of the classroom assessments and the high cost of communication systems to support the decentralize nature of the process. However, in order to have a sound system that will meet NCLB reporting requirements, these components are essential. So, although they are costly, they are unavoidable expenses to support this classroom-based system.

Conclusion

This report describes a classroom-based science assessment system that would meet the requirements for the No Child Left Behind Legislation. In particular, the statewide science assessment system must:

- Measure the state's challenging academic standards in science.
- Be aligned with the state's science standards and involve multiple up-to-date measures of student academic achievement.
- Include either (or both) criterion-referenced assessments or augmented norm-referenced assessments.
- Specify at least 3 achievement levels.
- Measure the achievement of all children and provide for participation of all students. Reasonable adaptations or accommodations should be made for students with disabilities and limited English proficient students.
- Report results in aggregate for the full group of test takers, disaggregated for specified population groups, and at the individual level. Reports should include both descriptive and diagnostic information.

The proposed classroom-based science assessment system can, if implemented as outlined in this document and if sufficient professional development is provided, meet all of these requirements. It provides for specialized professional development to aid teachers in their understanding of and use of the state's science content standards. One of the technical quality indicators on which the classroom assessment will be evaluated is the alignment of the assessments to the state's science standards. Further, by virtue of using the classroom as the site for the administration of the

assessments, which will be included in the instructional unit, teachers will have more flexibility in the kinds of assessment tasks they employ and have multiple opportunities to observe their students' learning and achievement on the content standards. The assessments included in the classroom-based model are criterion-referenced by design. However, collateral information provided by other norm or criterion-referenced assessments will be valuable in providing supportive evidence for the veracity of the results from the classroom-based science assessment system. The system is flexible enough to accommodate at least 3 achievement levels, however, the use of more levels puts increasing pressure on the teachers to ensure that their assessments have sufficient breadth and depth to make these multiple classification decisions. The same assessment system will be used uniformly across the state, as it is embedded in regular classroom activities. Because English language learners and special needs students will have the opportunity to participate in the assessment program in ways that are consistent with their typical classroom experiences, there is reason to believe that this classroom-based system will be more appropriate for these students than would other more centralized assessment practices. Finally, because the results will be reported in pre-specified performance categories, aggregation of results across the state will be feasible. Assuming that the state will have an acceptable database structure, disaggregated results for specified groups will be possible. Reports can be tailored for different stakeholders and have the benefit of being available in a timely manner. Diagnostic information will be available at the classroom level to inform students, parents, and teachers.

Therefore, as proposed the model provides a coherent, instructionally useful, teacher-led assessment program that meets the various criteria for technical rigor and that can provide adequate reporting at different levels and to different stakeholders. This is what was required for the classroom-based science assessment system by the NRC Committee on Test Design for K-12 Science Achievement.

References

Buckenhahl, C,W., Impara, J.C., & Plake, B.S. (2002). District accountability without a state assessment: A proposed model. Educational Measurement: Issues and Practice, 21, 6 - 16.

Buckendahl, C.W., Plake, B.S., & Impara, J.C. (2004). Educational Measurement: Issues and Practice, 23.17 – 25.

No Child Left Behind Act of 2001. (Pub. L. No. 107-110, 115 Stat. 1425 (2002).

Plake, B.S., Buckendahl, C.W., & Impara, J.C. (2004). Educational Measurement: Issues and Practice, 23.12 - 16

Roschewski, P., Isernhagen, J., & Gallagher, C. (April, 2001). The Nebraskans reach for the STARS. Phi Delta Kappan. 661-615,

Roschewski, P.(March, 2003). Nebraska STARS line up. Phi Delta Kappan, 517-520.

Appendix

Appendix A: Shorts Biographies of Buros Team Members

Buros Center for Testing Classroom-based Science System Models Committee

Short Biographies of Committee Members

Barbara S. Plake, Ph.D., University of Nebraska-Lincoln; W.C. Meierhenry Distinguished University Professor, Director of the Oscar and Luella Buros Center for Testing, and Director of the Buros Institute of Mental Measurements. Dr. Plake joined the UNL faculty in 1978 after receiving her Ph.D. in Educational Statistics and Measurement from the University of Iowa and working at American College Testing Programs. She has served the measurement community in several roles: by co-founding the scholarly journal Applied Measurement in Education, serving on the Board of Directors of the National Council on Measurement in Education (NCME), and serving as President of NCME in 1992-93. She has authored over 100 refereed publications and serves in an advisory capacity to many educational agencies and professional associations. Her expertise is primarily in the areas of teacher assessment literacy, state assessment and accountability, computerized testing, including adaptive testing methods, and licensure/certification testing, including setting of performance standards or cutscores. She has served as a consultant to Nebraska, Massachusetts, Virginia, Missouri, South Dakota, and Connecticut Departments of Education for their State assessment programs.

Chad W. Buckendahl, Ph.D., University of Nebraska-Lincoln; Director of the Buros Institute for Assessment Consultation and Outreach, a division of the Oscar and Luella Buros Center for Testing. Dr. Buckendahl has worked with Buros since 1998 and received his Ph.D. in Quantitative and Qualitative Methods in Education from the University of Nebraska – Lincoln in 2000. Before coming to Buros, he worked with The Gallup Organization. His expertise is primarily in the areas of state assessment and accountability, standard setting, and licensure/certification testing. He has consulted with educational testing programs in Nebraska, South Carolina, and South Dakota; licensure testing programs in education, dentistry, and law; and assisted in the development of an accreditation program for proprietary tests.

James C. Impara, Ph.D. Buros Center for Testing; MA., Ph.D Florida State University. Impara has spent 33 years in the testing field including designing and managing statewide testing programs. During his 20-year tenure in higher education, he consulted with states, business, and licensure and certification agency testing programs. He served as Director of the Buros Institute of Assessment Consultation and Outreach, from 1995 to 2003, where he worked with various testing clients. He has performed audits and accreditations of educational, licensure, and certification programs. Impara has also participated in the development of testing program quality standards for the Buros Institute, the Conference for Food Protection, and the National Association of State Contractor Agencies. Although Impara has conducted testing program audits that look at all aspects of testing, he specializes in conducting security audits, preparing security plans and providing diagnoses of test security reports. Impara has chaired and served on a variety of different committees for the American Educational Research Association and the National Council

on Measurement in Education where he served a three-year term as a board member and is current President-elect. He continues to be a reviewer for several measurement journals, while co-editing *Applied Measurement in Education*. He is also a co-editor of the *Mental Measurements Yearbook* Series.

Arthur I. Zygielbaum is Co-Director of the National Center for Information Technology in Education (NCITE) which is part of the College of Education and Human Sciences at the University of Nebraska – Lincoln (UNL). He also holds a courtesy appointment as Associate Professor of Computer Science and Engineering. Mr. Zygielbaum is leading efforts to discover how to effectively apply technology to improve teaching and learning. A past member of the NASA Aerospace Safety Advisory Panel, he is also an aerospace consultant. He is currently on an independent review team studying safety and engineering management in the US Navy. Art Zygielbaum joined UNL in January of 1998 after a nearly 30-year career as an engineer and manager at the NASA/CALTECH Jet Propulsion Laboratory (JPL). He is an instrument-rated private pilot and amateur radio operator. He received his Bachelors degree in Physics from the University of California, Los Angeles, and a Masters in Electrical Engineering Computers from the University of Southern California. As time allows, he is a graduate student in Educational Psychology at UNL.

Scott Marion, Ph.D. is a Senior Associate at the National Center for the Improvement in Educational Assessment. He is a recognized national leader in designing statewide accountability systems under No Child Left Behind and was the co-chair a joint study group project for CCSSO's Accountability and Comprehensive Assessment System SCASS projects focusing on the implementation of No Child Left Behind. Prior to joining the Center for Assessment, Scott was most recently the Director of Assessment and Accountability for the Wyoming Department of Education and was responsible for overseeing the Wyoming Comprehensive Assessment System (WyCAS), Title I assessment and accountability, assessment requirements related to Wyoming's district accreditation model, and the Body of Evidence assessment system for high school graduation. A former biologist and high school science teacher with a Master's in Science Education from the University of Maine, Scott originally moved west to complete his Ph.D. at the University of Colorado, Boulder in educational measurement under the tutelage of Lorrie Shepard and Robert Linn. At CU, Scott had the good fortune of collaborating with several faculty members, including Hilda Borko, Margaret Eisenhart, Ernie House, and Lorrie Shepard, on research and evaluation projects leading to peer reviewed publications and technical reports.

Ian D. MacGregor, Ph.D.. MacGregor was educated at Aberdeen University, Scotland (B.Sc. Hons, 1957), Queen's University, Canada (M.Sc., 1960), and Princeton University, NJ (Ph.D., 1964). During his early career he worked as a field geologist with the Geological Survey of Canada, followed by a post doctorate research experience at the Geophysical Laboratory, Carnegie Institution of Washington. He has been a professor, at the Southwest Center of Advanced Studies (now the University of Texas at Dallas) and the University of California at Davis. In addition, he worked at the National Science

Foundation where he served as a Section Head for Major Projects and Division Director of the Earth Science Division. In the last three years he has worked with colleagues at the national Science Resources Center learning about K-12 education and has worked on the development of Middle School science curricula. Most important has been the opportunity to learn how an individual with skills in scientific research and university teaching may make a contribution to K-12 science education.

Cindy L. Gray, Ph.D., Director of Curriculum and Assessment Elkhorn Public Schools, Elkhorn, NE; Dr. Gray joined Elkhorn Public Schools after receiving her Ph.D. in School Psychology from the University of Nebraska – Lincoln in 2000. Prior to her work in Elkhorn, she served as a middle school English teacher, a gifted education facilitator, a high school guidance counselor, and an intern school psychologist. In addition to work with the public schools, she is also an instructor for the University of Nebraska- Lincoln Assessment Cohort, an 18 credit hour graduate program designed to develop assessment literacy and leadership skills among practicing teachers and administrators. Dr. Gray serves on a variety of state committees. She is president-elect for the Nebraska Association for Curriculum and Supervision Development, a member of the Nebraska Educational Leadership Institute, a participant in the Nebraska Department of Education Adequate Yearly Progress Task Force, and a member of the Nebraska Essential Education Policy Study Committee.

Teresa J. Eckhout, M.A., Elementary Assessment Coordinator, Lincoln Public Schools. Ms. Eckhout accepted a teaching position with Lincoln Public Schools in 1991, after receiving her bachelors degree in Elementary Education from the University of Nebraska – Lincoln in 1990 and teaching in a local parochial school. In 1999, she was hired by the Educational Service Unit in Lincoln Public Schools to develop district level assessments for reporting to the Nebraska Department of Education on student achievement in the areas of math, reading/language arts, and science. She has supported district efforts in the areas of assessment literacy, assessment development, and alignment of district objectives and classroom assessments. She currently serves as co-investigator for the Improving Teacher Quality: State Grant Program, Inservice and Preservice Assessment Literacy Study Groups (IPALS) project. In 2003, she received a masters degree in Educational Psychology, Quantitative and Qualitative Methods in Education from the University of Nebraska – Lincoln. Ms. Eckhout has recently accepted a position as Project Coordinator for the Buros Institute for Assessment Consultation and Outreach.

John McKinney McKinney is a science teacher at Mountain Ridge Middle School in Douglas County, Colorado. He is also an adjunct Professor of Education at Adams State College, teaching graduate courses in scientific inquiry and the application of educational research in classroom instruction. His expertise is in the area of creating science assessments at the classroom level. He is recognized for creating and implementing science standards in his district, and consulted on the creation of the state science test. He speaks regularly on developing classroom assessment systems that support student learning and assess higher level thinking. John holds a masters degree in Earth Sciences from the University of Northern Colorado. His

teaching at the middle level has earned him the Milken Educator award and the National Earth Science Teacher of the Year award.

Thomas A. Neumann, Masters Degree in Secondary Science Education-1990, Nebraska Assessment Cohort-2001, Certified in Classroom Assessment Practices-2003, Physics/AP Physics Teacher Millard Public Schools-Millard South High School. Recipient of the 2004 Radio Shack National Teacher Award.

Carol Shestok . K-5 Science Coordinator and Mentor Teacher Training Coordinator and Lead Trainer for the Westford Public School System, Westford, MA. She is National Board Certified Teacher (NBCT). She was one of the first 22 NBCTs in the Commonwealth of Massachusetts. Carol has been teaching 25 years and received her B.S. and M.Ed. from Kutztown University, Kutztown, PA. She is currently in the Doctoral Program at the University of Massachusetts Lowell. She is also working on the principal's certification. She has taught in Massachusetts, Pennsylvania, and Connecticut. Carol has presented at various state, regional and national conferences. In addition, as a member of the NBCT Standards Delegation, she presented at universities in China and Australia. Other awards she received include: Presidential Award for Excellence in Mathematics and Science Teaching; Massachusetts Teacher of the Year Finalist, 2000; Town of Westford Seavey Award for Community Service, 2003; EPA Environmental Educator; MA Secretary's Award-Excellence in Environmental Education; New England Water Works Teacher Award; Merrimack Education Collaborative Instruction Award and Peter Farrelly Award-Teaching Excellence. She has published works in a college pre-service text and state and national periodicals. She is a member of Phi Lambda Theta and Kappa Delta Pi and a Fellow in the Massachusetts Teacher Leadership Academy. She also is an adjunct instructor for Fitchburg State College, Fitchburg, MA.

J. Myron Atkin, Ph.D. Professor of Education (Emeritus), Stanford University
Mike Atkin taught science for seven years in New York elementary and secondary schools. He joined the faculty of the University of Illinois at Urbana-Champaign in science education in 1955 and moved to the Stanford University faculty in 1979. At both universities, he also was Dean of Education -- from 1970 to 1979 at Illinois and from 1979 to 1986 at Stanford. He has chaired the Education Section of the American Association for the Advancement of Science and served as a consultant on education to the Organization for Economic Cooperation and Development (OECD) in Paris. He is a National Associate of the National Academies of Science/National Research Council, where he was a member of the Mathematical Sciences Education Board, the National Committee on Science Education Standards and Assessment, and chair of the Committee on Science Education K-12. He chaired the Committee on Science and Engineering Education of Sigma Xi, the honorary scientific research society, and was vice-chair (1985-86) of the Advisory Committee for Science and Engineering Education at the National Science Foundation. During 1986-87, he served as Senior Advisor to the Education Directorate at NSF. In the 1960s, he directed one of

the first two NSF-supported curriculum projects for children below the high school level, the University of Illinois Astronomy Project.

Technology Advisory Committee

Roger H. Bruning, Ph.D.Bruning is Velma Warren Hodder Professor of Educational Psychology and Co-Director of the Center for Instructional Innovation and the National Center for Information Technology in Education at the University of Nebraska-Lincoln (UNL). A Nebraska native, Dr. Bruning received his Bachelor's degree in Modern Languages and Literatures from the University of Nebraska in 1963. He completed his M.A. in Educational Psychology in 1965 and his Ph.D. in Educational Psychology in 1968.
Since joining the faculty of the University of Nebraska, Dr. Bruning has taught courses in applied cognitive psychology, cognition and technology, literacy, and educational measurement and research. During that time, he has written numerous articles for journals such as the Journal of Educational Psychology and Scientific Studies of Reading. He has served as a consultant for many education- and health-related projects and as evaluator for projects of the National Science Foundation, Howard Hughes Medical Institute, Agency for International Development, and the Environmental Protection Agency. He is a Fellow of the American Psychological Association (Division 15) and has served on the Editorial Boards of the Journal of Educational Psychology, Educational Psychology Review, Contemporary Educational Psychology, and Scientific Studies of Reading.

Gwen Nugent, Ph.D. Nugent is Research Associate Professor of the National Center for Information Technology in Education and Director of Education at Nebraska Educational Telecommunications, both at the University of Nebraska-Lincoln. Dr. Nugent coordinates development and research projects focusing on the impact of technology to improve student learning and teacher competencies, with special emphasis on multimedia instruction and on-line assessment. She has over 30 years experience in the design, production, and evaluation of mediated instruction, and has served as project manager for over 300 multimedia projects designed in a variety of subject areas and for a variety of audiences. The majority of these projects are distributed nationally and internationally, and many have won national awards for their educational impact and effectiveness.

Al Steckelberg, Ph.D Allen L. Steckelberg is an assistant professor, in the Department of Teaching, Learning, and Teacher Education in the College of Education and Human Sciences at the University of Nebraska-Lincoln. Dr. Steckelberg coordinates graduate and undergraduate instructional technology programs within the College. He received his Masters from the University of Nebraska in 1978 in Special Education and his Ph.D. from the University of Nebraska, in 1992 in the area of Psychological and Cultural Studies. Areas of teaching and research include teacher education, technology in

education, Web-based instructional and educational management, and paraprofessionals in school programs. He has served as director, PI or co-PI on 16 federally sponsored grants representing $ 4.9 million in external funding. These projects have produced a number of widely used Web-based training and educational management resources.