EDUCATIONAL RESEARCH ABSTRACTS

G. BASS, Jr.- Section Editor
The College of William and Mary
Williamsburg, VA 23187

***Editor's Note**: As noted in the first issue of The Journal of Mathematics and Science: Collaborative Explorations, the purpose of this Educational Research Abstract section is to present current published research on issues relevant to math and science teaching at both the K-12 and college levels. Because educational research articles are published in so many different academic journals, it is a rare public school teacher or college professor who reads all the recent published reports on a particular instructional technique or curricular advancement. Indeed, the uniqueness of various pedagogical strategies has been tacitly acknowledged by the creation of individual journals dedicated to teaching in a specific discipline. Yet many of the insights gained in teaching certain physics concepts, biological principles, or computer science algorithms can have generalizability and value for those teaching in other fields or with different types of students.*

In this review the focus is on "assessment." Abstracts are presented according to a question examined in the published articles. Hopefully, such a format will trigger your reflections about exemplary math/science assessment as well as generate ideas about your own teaching situation. The abstracts presented here are not intended to be exhaustive, but rather a representative sampling of recent journal articles. Please feel free to identify other useful research articles on a particular theme or to suggest future teaching themes to be examined. Please send your comments and ideas via email to gmbass@facstaff.wm.edu *or by regular mail to The College of William and Mary, P. O.Box 8795, Williamsburg, VA 23187-8795.*

Assessing Student Performance in Mathematics and Science

> "If tests determine what teachers actually teach and what students will study for - and they do - then test those capabilities and habits we think are essential, and test them in context."
>
> Grant Wiggins (1989), President and Director
> Center on Learning, Assessment, and School Structure

The Journal of Mathematics and Science: Collaborative Explorations Volume 2 No 1 (1999) 105-114

What is the purpose of testing students? Should tests serve an auditing function that simply provides a specific score for each student's performance? Should tests be the chief incentive that motivates student effort? Should tests be the key mechanism through which teachers determine merit ratings and placement decisions for individual students? Should tests be one component of a system that improves learning and instruction? Wiggins clearly believes testing is only one part of the larger issue of assessment and that the primary purpose of student assessment is for students to learn better and for teachers to teach better.

The National Council of Teachers of Mathematics published in 1989 their influential Curriculum and Evaluation Standards for School Mathematics. They categorized their fourteen evaluation standards according to a focus on general assessment, student assessment, and program evaluation. Four key themes underlie the 1989 NCTM evaluation standards:

- student assessment be integral to instruction;
- multiple means of assessment methods be used;
- all aspects of mathematical knowledge and its connections be assessed; and
- instruction and curriculum be considered equally in judging the quality of a program

Very practical classroom applications to assess what student know and how they think about mathematics were identified in that document. The Standards advance more attention to such assessment strategies as taking a holistic view of mathematics, developing problem situations that require the application of a number of mathematical ideas, and using standardized achievement tests as only one of many assessment indicators. (The 1989 NCTM Standards can be accessed on the Web at *http://standards-e.nctm.org/1.0/89ces/Table_of_Contents.html*)

Recently, the National Council of Teachers of Mathematics released a draft copy of their updated "Principles and Standards for School Mathematics." (The Standards 2000 project report can be found at *http://www.nctm.org/standards2000/*). These standards for the 21st century continue the emphasis on good assessment practices. The Assessment Principle guiding the new NCTM recommendations states "Mathematics instructional programs should include assessment to monitor, enhance, and evaluate the mathematics learning of all students and to inform teaching." In their continued emphasis on assessment as a process, the Standards 2000 draft asserts that teachers need to utilize a classroom assessment cycle that involves four key activities: Setting clear goals (planning assessment); Gathering evidence using various methods; Interpreting evidence (making inferences); and Making decisions (taking action.) In the web-version of the Standards, electronic examples of how teachers can practice "best assessment" in each part of the cycle are illustrated.

In 1991 the National Research Council initiated an ambitious effort to develop national standards for science education in content, teaching, and assessment. They proposed five assessment standards:

A: Assessments must be consistent with the decisions they are designed to inform;

B: Achievement and opportunity to learn science must be assessed;

C: The technical quality of the data collected is well matched to the decisions and actions taken on the basis of their interpretation;

D: Assessment practices must be fair;

E: The inferences made from assessments about student achievement and opportunity to learn must be sound.

However, it is in Teaching Standard C that the National Science Standards advocate "best practice" for teachers' classroom assessment:

<u>TEACHING STANDARD C:</u>

Teachers of science engage in ongoing assessment of their teaching and of student learning. In doing this, teachers

- Use multiple methods and systematically gather data about student understanding and ability.
- Analyze assessment data to guide teaching.
- Guide students in self-assessment.
- Use student data, observations of teaching, and interactions with colleagues to reflect on and improve teaching practice.
- Use student data, observations of teaching, and interactions with colleagues to report student achievement and opportunities tolearn to students, teachers, parents, policymakers, and the general public.

Clearly, student assessment is a very important part of educational reform in both mathematics and science. Amid the national claims and counterclaims as to the best way to conduct student assessment, it is often difficult for a K-12 teacher or college professor to know what exactly what kind of testing to do. In that context it is valuable to remember what

Elizabeth Badger, the Director of Assessment for the Massachusetts Department of Education, said in 1992,

> "A perfect test or perfect task does not exist.... Almost any task can be used, provided that we recognize what information we want to obtain from it."

How do teachers from elementary school through university level assess student achievement in mathematics and science? Are current practices the best ways to assess student understanding in mathematics and science? What alternative assessments techniques are being promoted and pilot-tested in mathematics and science classrooms? What valuable lessons have been learned from classroom research? What recommendations do K-12 and college-level instructors publicly offer about student assessment? The following set of articles provides a sample of recent academic writings on the subject of student assessment.

• How can science professors make course examinations more creative, more meaningful, and more useful?

"Every faculty member knows that exams drive student behavior." So begins the preface to the 1997 book *The Hidden Curriculum - Faculty-made Tests in Science* by Sheila Tobias and Jacqueline Raphael. A year earlier they used the same quote to begin this article exploring how new theories about testing might lead college professors to new assessment practices. Findings from cognitive science research investigating expert-novice differences and the subject-specific heuristics in various disciplines have led college faculty to reconsider their teaching goals, classroom practices, and student testing procedures. For example, the Force Concept Inventory developed by physics educators David Hestenes and Ibrahim Halloun has verified that passing a college physics course does not eliminate many students' misconceptions about force, mass, acceleration, and mechanics. Increasingly, other science professors have recognized that tests emphasizing algorithmic problem-solving strategies ("plug-and-chug") will not necessarily encourage nor assess the students' conceptual understanding of scientific principles.

Tobias and Raphael lament that most individual professors' in-class testing innovations never become widely known among their colleagues. Their study of college science professors was intended to remedy this unfortunate circumstance. They collected 160 narratives of college science faculty's efforts to integrate course curricular goals with new examination strategies. In this article they highlight a sample of innovative science testing while a fuller description is available in their book. Among the new assessment practices they identify:

- Portfolio assessment by an astronomy professor
- Multi-step, "real-world" chemistry problems by an introductory chemistry professor
- Group format midterm exams by an earth science professor
- Optional grade-performance contracts by an organic chemistry professor
- Grading students' work for both content and coherence, i.e. algorithmic solutions and conceptual understanding, by a physics professor

In separate focus group interviews with undergraduate and graduate students, Tobias and Raphael examined what both science majors and non-science majors thought of their science exams. Short answer and essay questions were preferred to multiple choice items. Grading students' performance "on the curve" was felt to introduce competition and relative comparisons that students felt often masked knowledge gained through hard work. Timed tests were also seen as adding unnecessary stress to an already tense test-taking situation. In general, students wanted a diversity of assessment strategies used in their science courses, because of the diversity of learning styles among the students. In this article Tobias and Raphael introduce a variety of testing strategies used in actual college science classrooms and point the interested reader to a fuller description in *The Hidden Curriculum.*

S. Tobias and J. B. Raphael, "In-class examinations in college-level science: New theory, new practice", *Journal of Science Education and Technology*, 5 (4), 311-319 (1996).

• How can mathematics teachers improve their classroom test items?

In the 1995 Assessment Standards for School Mathematics, the National Council of Teachers of Mathematics recommends teachers use multiple and complex assessment tools to judge "each student's attainment of mathematical power." The NCTM report advocates the increased use of performance tasks, projects, writing tasks, oral presentations, and portfolios and the decreased use of chapter quizzes and tests. Denise Thompson, Charlene Beckmann and Sharon Senk do not challenge the value of that recommendation, but they do contend there are strong practical reasons why classroom math teachers will continue to use in-class tests. They do propose ways to analyze and modify typical classroom tests that are reasonably objective, easy to implement, and consistent with NCTM recommendations.

Thompson and her colleagues have created a test item classification scheme with which teachers can examine their classroom tests. Eight categories (*item format; level; skill; real context; open-ended; graphical representation; reasoning; technology*) can be used to analyze and then modify test questions to reflect the most important aspects of the mathematical curriculum. For example, *level* helps the teacher differentiate prerequisite skill items that require one or two steps to complete from other items requiring multiple steps. The *graphical representation* category reminds the teacher to identify test items that (1) entail graphical interpretation to find the answer, or (2) require a graphical construction as the answer, or (3) provide a graph that is superfluous for finding the answer, or (4) provide no graph or diagram. The *technology* category helps identify items that require a tool, permit a tool, or exclude the use of a tool. They use test items in algebra, geometry, and precalculus to illustrate how their approach can indeed produce exams that align with the NCTM mathematical content and processes. As Thompson, Beckmann, and Senk point out, "Students tend to value what is graded." Therefore, graded class tests should reflect what important concepts and skills teachers indeed value.

D. R. Thompson, C. Beckmann, and S. L. Senk, "Improving classroom tests as a means of improving assessment", *The Mathematics Teacher*, **90** (1), 58-63, (1997).

- **How can a professor use a student's misunderstanding on a class exam to increase physics learning?**

Any teacher who has ever given an exam knows there is often a wide gap in what students think they have learned and what teachers think they have taught. In fact, a course exam lets both student and teacher get a more public confirmation of what was learned (and not learned!). Unfortunately, some students only realize their misunderstanding after they have been confronted with a test problem they fail to complete successfully. Is it too late to use such "painful enlightenment" as an integral part of the teaching/learning process? While an exam should be a reliable and valid measure of student understanding, it can also be an assessment that guides future learning.

Thomas Ammirati, a physics professor, asks the intriguing question "Why not treat a student's test performance as a work in progress?" In recent years his students had expressed their frustration with physics exams that resulted in a low class average, e.g. below 70. Ammirati recognizes that his discouraged students might explain the overall poor class performance to unrealistically difficult test questions, to students' lack of ability and/or effort or, worse yet in his mind, to overall poor teaching. To remedy both his students' and his own frustration, he began offering students the option of reworking exam problems they originally had not completed successfully. He allows students to redo selected problems under "open book" circumstances, to meet with him individually for a review of this revised work, and to submit that work for him to regrade. Students can earn up to 50% of the lost points for each problem they successfully revise. Ammirati sees this targeted test revision opportunity allowing students to clarify and relearn misunderstood physics principles as well as add points to their test grades. He acknowledges that his conferencing with students is time intensive, but he has not had to add additional office hours. Now he really has students use his normal office hours to meet with him. His apprehensions that students might take the original exam less seriously has not happened since students are never sure which test items they will have the opportunity to revise. Finally, he believes students work toward a better understanding of physics through these problem revisions because they also know there is no revision option on the final exam. His students are better prepared for the final because they have a better understanding of the physics material covered in the course.

T. Ammirati, "Targeted test revision- another approach to science testing", *Journal of College Science Teaching*, **28** (2), 117-120 (1998).

- **How can testing in large lecture course actually lead to more interactive student learning?**

Since so many introductory college science courses are taught in large lecture settings, beginning college science students must not only master the challenging course content, but do it in a format that makes two-way communication with the instructor difficult. Course exams do provide students feedback about tested concepts, but seldom are given frequently enough to compensate for the large class size. In his chemistry course (83 students enrolled), Thomas Holme regularly uses a technique he calls Interactive Anonymous Quizzes. He typically gives students a few multiple-choice questions that focus on key chemistry concepts at the beginning of the class. Students are given a few minutes to select the correct answer and to indicate their level of confidence ("Very confident," "Somewhat confident," "Just guessing"). Students are next given three minutes to convince those seated near them that they are correct. After these peer discussions, the students are again asked to select the alternative they believe is correct and indicate their confidence level. The students' quizzes are collected anonymously at the end of the class at which time the professor provides the correct answer.

Holme believes this simple technique helps both the professor and students. The teacher is able to scan the student answers very quickly to see if the most of the class understands each item. For questions with poor student performance, the professor can plan how to clarify or reteach key principles. The students are able to get almost immediate feedback on their chemistry understanding, first from other students and eventually from the professor.

Holme has used Interactive Anonymous Quizzes in classes as large as 250 students. He observes that only rarely do students change from the correct answer to an incorrect answer. Students regularly express more confidence in their answers after talking with peers even if they originally selected the correct answer. In most cases a sizable percentage of students (20% to 34%) switch from the wrong answer to the correct one on each item. Holme concludes these quizzes provide improved communication and student interaction which results in greater learning. Holme further reasons that the short amount of class time spent on these quizzes and relatively nonintrusive way it can be incorporated in a large lecture course makes Interactive Anonymous Quizzes an effective testing-teaching technique.

T. Holme, "Using interactive anonymous quizzes in large general chemistry lecture courses", *Journal of Chemical Education*, **75** (5), 574-576 (1998).

Additional Resources on Assessment Research and Practice
Some helpful electronic resources to explore:

Balanced Assessment in Mathematics is a National Science Foundation project charged with

developing new approaches to the assessment of mathematical competence in the elementary and secondary grades.
http://edetc1.harvard.edu/ba/

The ERIC®Clearinghouse on Assessment and Evaluation seeks to provide 1) balanced information concerning educational assessment and 2) resources to encourage responsible test use.
http://ericae.net/MAIN.HTM

Assessment & Evaluation on the Internet is a very complete list of links to other websites and documents on 40 topics from Action Research to Tests Online including mathematics and science assessment.
http://ericae.net/intbod.stm

Pathways to School Improvement
This website is a product of the North Central Regional Educational Laboratory in cooperation with the Regional Educational Laboratory Network and provides research-based resources and assistance to educators on a wide variety of topics including evaluation and assessment.
http://www.ncrel.org/ncrel/ncrel/ncrel/sdrs/areas/as0cont.htm

The Eisenhower National Clearinghouse for Mathematics and Science Education provides educational journal articles, teaching programs, and educational standards on the latest teaching trends and developments including assessment. For example, "Assessment in Action," a collaborative action research report focused on mathematics and science assessments, reports of twenty-three teacher research projects is found on this site.
http://www.enc.org/

Developing Educational Standards is an annotated list of Internet sites with K-12 educational standards and curriculum frameworks documents, maintained by Charles Hill and the Putnam Valley Schools in New York.
http://putwest.boces.org/StSu/Science.html

Some useful books on assessment to examine:

Mathematical Sciences Education Board, *Measuring what counts*, National Academy Press, Washington, DC, 1993.

G. Phye, *Handbook of classroom assessment*, Academic Press, San Diego, CA, 1997.

W. Popham, *Classroom assessment, what teachers need to know,* Allyn & Bacon, Needham Heights, MA, 1995.

T. Romberg, *Reform in school mathematics and authentic assessment*, State University of New York Press, Albany, NY, 1995.

S. Tobias and J. Raphael, *The hidden curriculum part I,* Plenum Press, New York, 1997.

S. Tobias and J. Raphael *The hidden curriculum* part *II*, Plenum Press, New York, 1997.

G. Wiggins, *Educative assessment, designing assessments to inform and improve student performance*, Jossey-Bass Publishers, San Francisco, 1998.