A New Direction for Regents Examinations in Science Education

from the January 2004 issue of ***Perspectives in Science Education***

Paul Zachos
Association for the Cooperative Advancement of Science and Education
110 Spring St., Saratoga Springs, NY 12866, phone: 518-583-4645, fax: 518-587-6467
paz@acase.org

Non-Educational Use of Tests.

As we enter the 21st century the increasing proliferation of tests which are used for *non-educational* purposes has become the most serious problem facing the field of educational assessment. How can a test be *non-educational*? The answer to this question requires a look at the heart of the educational process. There, we find invisible processes taking place deep within students that we call *learning*. We also find observable processes under the direction of teachers that we call *teaching* or *instruction*.

The purpose of teaching is to achieve goals for learning; this is the one and only truly educational function. In practical day-to-day teaching, goals for learning are expressed as *learning objectives*.

Educational assessment is the process of obtaining information that serves as evidence that *learning objectives* have been attained. Assessment information may be derived from homework, lab reports, tests, essays, and even classroom conversations.

The educational value of assessment information lies in its appropriateness and availability to be useful in planning and improving teaching. Use of assessment information for other purposes (e.g. grading) is *non-educational*! When we use the information about students' levels of attainment on *learning objectives* to decide what we will teach and how we will teach it, we are serving a direct educational purpose; our assessment results are being used to support teaching and learning. When we add up the correct items on a test and assign a score or letter grade, no matter how noble our intentions are, no matter how carefully we think through our formulas, **we are not serving a direct educational purpose because these numbers are not directed to achieving *learning objectives*.**

The Critical Role of Learning Objectives for Assessment and Instruction.

Valued learning objectives refer to concepts, skills, or dispositions that are productive for future learning, vocation, and community life. *Valued learning objectives* are typically derived from State and National Standards, curriculum documents, and recommendations of science advisory councils (e.g. (AAAS Benchmarks, 1993)). They can also be identified empirically (Zachos, Hick, Doane and Sargent, 2000). A test designed to directly support teaching and learning would be based on *valued learning objectives*. It would give teachers a report of the competence of each student on each *learning*

objective associated with the targeted curriculum unit. The teacher could then use this information to plan instruction in order to maximize student attainment of the *objectives*.

Are Our Tests Educational?

Think back and reflect. How many of the tests that we administer provide information we will use to plan and improve instruction for the group of students who were tested? Typically, tests are given when instruction is complete. In this situation, there is no chance for the test results to support teaching and learning. At best, instruction can be improved for future classes. Post-instruction tests may support some learning if test results are accompanied by feedback to students, but cannot influence instruction because instruction is over. The prime time for educational assessment is before or during, not after, the process of instruction. Moreover, it is very difficult for tests (teacher-made and/or standardized high stakes tests) to be useful for planning instruction when they do not provide information about whether specific *learning objectives* were attained. Consequently, instead of having tools with which to maximize learning, we typically end up with scores that rank students on the number of test items they answered correctly. **The fact that a test is administered in school does not mean that it is serving an educational purpose.**

Is a New Direction Possible?

Is instruction based on assessment unrealistic? Not at all! Methods for setting *learning objectives* and building *performance assessments* which measure them were developed in the last half of the 20th century (Mager, 1962; Zachos et al., ibid.). Ways to administer such assessments efficiently in classroom settings are also available (Zachos, in press). Teachers can learn to make reliable judgments on student performance (Brennan, 2001). Information technologies for keeping track of student performance on important *objectives* across time (e.g. across the secondary school career) are now available and practical for use in school settings. Problems of test security can be shown to be more easily solvable when assessing higher order cognitive abilities (e.g. upper levels of Bloom's taxonomy, 1956) than the type of knowledge assessed by forced choice test items. The pieces needed for efficient *objectives-based assessment* exist and await integration into existing practices. Many past attempts to integrate the pieces have failed because of overly ambitious expectations for teacher competence in building and using assessments, and because insufficient resources were allocated to assessment activities and professional development (Stiggins, 2001).

Two Scenarios – One Real and Undesirable, the Other An Unrealized Possibility.

Let's consider some realities and possibilities using the attainment of the concept of density as an example. Without question, an understanding of density and the ability to apply this understanding are fundamental to scientific literacy. Moreover, competence in density is needed for success in all high school science education programs. Consider two scenarios:

Scenario 1 – Real and Undesirable: *Earth Science teachers in a local school district are meeting to plan the next year's curriculum. The recently administered Regents Examination had less emphasis on density than exams of previous years. The teachers*

consider that this may reflect a trend for future tests and deliberate on whether to reduce the instructional time devoted to density in the coming year. This, in spite of the fact that they believe that most of their students are leaving the course without an adequate grasp of the concept of density.

In this scenario, teachers must second-guess the test design process because the tests are not explicitly aligned to *learning objectives*. Moreover, in the end, they are left with *beliefs* and not concrete knowledge concerning student competence related to density. In fact, there typically is no information available about the extent to which students have or have not attained competence in using and applying the concept of density at the class, school, or state level.

Scenario 2 – An Unrealized Possibility: *Regents examinations begin to be built around defined sets of valued learning objectives including, for example, the application of the concept of density in practical situations. Examination results are followed by reports of the level of attainment of these learning objectives. Teaching in every content unit is directed to the same learning objectives. Teachers have available to them assessment techniques that allow them to monitor student progress on the targeted objectives over the course of the school year. There are no surprises when the Regents Examination results arrive. The teacher has been intimately aware of levels of student competence on the learning objectives all along.*

This scenario is associated with an unexpected benefit. Because assessments are conducted at the level of the *learning objective*, the possibility of a unity of information emerges. The same information used by the teacher to plan and improve instruction (e.g. level of competence in applying the concept of density to real life problems) turns out to be the same information that the school district needs for public accountability. This same information is also what would be most salient to the State Education Department for monitoring progress at all levels, and for allocating resources (e.g. grants, research, professional development) to improve student performance. Currently, we have single scores on high stakes exams which misleadingly sum performances on diverse content areas and serve no educational purpose. Students are ranked by these scores and subjected to unproductive pass/fail criteria (Zawicki & Jabot, January 2003) based on scaled scoring grades set by arcane formulas.

The June 2003 New York State Regents Examination in Physics and the Need for a Paradigm Shift

Controversies associated with the June 2003 Regents examination highlight the problems identified above. Zawicki, Jabot, Falconer, MacIsaac, Henry and Fischer (October 2003) have conducted a carefully thought out and well-documented analysis of pass-fail performance on that examination. The analysis reveals the type of problems that we must face when our attention is captured by non-educational features of testing such as pass-fail levels. Contrary to commonly held beliefs, there is no scientific basis for setting a pass-fail level. Rather, a pass-fail level typically represents a value judgment which is realized as a mathematical distribution with certain ideal properties (e.g. a certain proportion of students passing and failing). Item difficulty becomes a matter of interest in its own right, because it supports the realization of this idealized mathematical distribution rather than providing an opportunity to analyze student error patterns,

establish their cause, and specify appropriate instructional treatments. Expectations for student attainment become lost in a flurry of technical manipulations to get the right number of students "passed" and "failed." These psychometric devices were developed for purposes other than supporting teaching and learning and it is time that we re-directed our attention to assessment practices that provide a practical basis for improving instruction and learning. The analysis that should occur after a Regents examination is one that addresses the question of which valued concepts and skills are being successfully attained and which are not. The answer to this question can be reasonably followed by strategic action at the state and local level to solve the problems of teachers and schools that are failing to achieve objectives at a desired level.

I recognize that these ideas and interpretations are radically different from much current practice, but there is also much in current practice that is unproductive and oppressive, much that we should begin to separate ourselves from. Those familiar with "backward design" (Wiggins & McTighe, 1998) will recognize, in what has been proposed here, the role of useful measures of student attainment in transforming educational programs and building quality control of instruction. *Objectives-based assessment* gives stakeholders in the educational process meaningful information for planning and decision-making. This information can be used to support teachers and realize high aspirations for students. By moving in this direction we can make a dramatic shift towards the development of *High Value*, as opposed to *High Stakes* exams (Shiland, 2001).

References.

American Association for the Advancement of Science (1993). *Benchmarks for Science Literacy*. New York: Oxford University Press.

Bloom, B.S. (Ed.), Engelhart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (1956). *Taxonomy of Educational Objectives – The Classification of Educational Goals*. New York: David McKay Company, Inc.

Brennan, R. (2001). An Essay on the History and Future of Reliability from the Perspective of Replications. *Journal of Educational Measurement* 38 (4), 295-318.

Mager, R. F. (1962). *Preparing Instructional Objectives*. Belmont, CA: Fearon Publishers.

Shiland, T. (2001). Changing High Stakes Exams To High Value Exams. *Perspectives on Science Education* (April). Penn Yan, NY: New York State Science Education Leadership Association.

Stiggins, R. (2001). The Unfulfilled Promise of Classroom Assessment. *Educational Measurement*: *Issues and Practice* 20 (3), 5-15.

Wiggins, G. & McTighe, J. (1998) *Understanding by Design*. Alexandria, VA. Association for Supervision and Curriculum Development.

Zachos, P. (in press). Pendulum phenomena and the assessment of scientific inquiry capabilities. *Science and Education* (special issue devoted to the *International Pendulum Project)*. Copies available from the author by request, paz@acase.org).

Zachos, P., Hick, T.L., Doane, W.E.J., & Sargent, C. (2000). Setting Theoretical and Empirical Foundations for Assessing Scientific Inquiry and Discovery in Educational Programs. *Journal of Research in Science Teaching* 37 (9), 938-962.

Zawicki, J. & Jabot, M. (2003). The June 2002 Regents Physics Exam: A Brief Analysis. *Perspectives on Science Education* (January). Penn Yan, NY: New York State Science Education Leadership Association.

Zawicki, J., Jabot, M., Falconer, K, MacIsaac, D. Henry, D., Fischer, R. (2003). A preliminary analysis of the June 2003 New York State Regents Examination in Physics. *Perspectives on Science Education* (April). Penn Yan, NY: New York State Science Education Leadership Association.