

Discovering the True Nature of Educational Assessment *Paul Zachos*

Published in the: *Research Bulletin*, 9(2), 7-12
Published by: The Research Institute for Waldorf Education

Testing and particularly standardized tests are increasingly identified as sources of a host of ills that afflict contemporary education. The ills have been widely catalogued and discussed (Kohn, 2000; Sacks, 1999). Testing and standardization are not, however, in themselves, the causes of these ills; to think so would be an error and a distraction, because the real problem is far deeper and more widespread. It lies in the *non-educational uses of tests*, an almost universal, unhealthy practice.

The key to solving many of the problems that we associate with testing is the adoption of sound testing practices, in particular the use of what I will call *educational assessment*. In addition, I believe that the adoption of sound assessment practices by Waldorf Educators can contribute to a renewal of mainstream education in the 21st century. In order to assume such a leadership role, we must first understand how tests are currently misused and how tests *could be* used to support teaching and learning.

The Proper Use of Tests

We teach because there is something that we wish our students to learn. That something can be seen as a *learning goal*. If we imagine the developing human being we will also imagine faculties or capabilities that we see as desired attributes of that being. These may be facility in reading musical notes, the ability to reason with proportions, or understanding how the political, economic, and cultural spheres interface. Such imaginings are a first step in formulating practical *learning goals* that can be used to give direction to instruction. Educational activities are justified because they contribute to the attainment of worthy goals for learning (Johnson, 1977). Similarly, only tests that contribute to the achievement of learning goals can truly be considered educational. *Educational assessment* is, and should exclusively be, the process of finding out the extent to which specified learning goals have been attained. When we use information about students' levels of success in attaining our goals to decide what we will teach and how we will teach it, we are serving a truly educational purpose; our assessment results are being used to support teaching and learning.

The educational value of assessment information lies just in the appropriateness and availability of that information for use in planning and improving instruction. Use of assessment information for other purposes (e.g., to decide whether to pass or fail a student) is necessarily a *non-educational* use of testing! When we evaluate a piece of student work by considering its desirable and undesirable features and assign a grade to that work, when we add up the number of items answered correctly on a test and give a score, no matter how noble our intentions are, no matter how carefully we have thought through our formulas, we are not serving a direct educational purpose because these grades are not directed to achieving learning goals. We may in some way be serving our

school, our children's parents, our need to control student behavior, or a college admission process, but we are not serving the process of learning.

Current testing practices usually don't provide information that can be used to plan and improve instruction for the group of students who are being tested. Typically, tests are given when instruction is complete, at the end of a unit, section, semester, or school year. These assessments are known as *summative assessments*. That is, they provide a summary of past learning. Summative test results, by their very nature, cannot be used to support teaching and learning for those being tested. At best, instruction can be improved for future classes. Post-instruction tests may support some learning if test results are provided to students with detailed feedback regarding specific learning goals, but they cannot influence instruction because instruction on the topic of interest has already come to an end.

The best time for educational assessment is before or during the process of instruction, while opportunities for instruction and learning still exist, not after instruction has ended. Such assessments are referred to as *formative assessments* in that they are conducted in the course of instruction and so have the possibility, indeed the role, of influencing the direction and nature of instruction. But simply conducting formative assessments, rather than summative assessments, is not sufficient to assure the productive role of assessment in education. Tests cannot be useful for planning instruction (whether they are teacher-made or are conventional, standardized, or high stakes tests) when they do not provide information about the extent to which specific *learning goals* were attained. Without clearly specified learning goals and a sense of where students are at with regard to those goals teachers do not have the basis for crafting instruction that fits the students' educational needs.

As a result, instead of having tools with which to maximize learning, we typically end up with scores that only serve to rank students based on the number of test items they answered correctly or how well we feel that they performed on some task. The fact that a test is administered in school does not mean that it is serving an educational purpose. On the contrary, reflection and observation on this issue will show that most of the testing that occurs in schools is non-educational!

An Example of Educational Assessment

What does truly educational assessment look like? For several years now, my colleagues and I at the Association for the Cooperative Advancement of Science and Education have been working on the broad goal of developing the creative scientific capacities of secondary school science students. Our program is directed to developing the capabilities of students to conduct scientific inquiry and discover scientific concepts through their own investigations. The program is founded on our belief that in establishing scientific concepts for themselves, students form a personal relationship both to nature and to the scientific search for truth. Specifically, we work to help students attain the abilities needed to rigorously observe and describe their observations, to reason with proportions, to organize and analyze information, to build and test concepts, and to realize the impact

of their own actions and thinking on their judgments and conclusions (Zachos, 2004; Zachos, Hick, Doane, & Sargent, 2000).

We do this by presenting students with phenomena that give opportunities to experience natural forces at work (for example, floating and sinking objects, patterns of shadows over the course of the day and the year, and the balancing of objects on fulcrums). We pose challenges that require students to observe, conceptualize, and bring order to their representations of the phenomenon. From the students' responses we can assess their level of competence with respect to a variety of *core capabilities* and *concepts*. For example, we are interested in students' predictions and reasoning concerning whether a cube will float or sink in a given liquid when a larger cube of the same weight has just been shown to sink in that liquid. Students are asked to predict whether the smaller cube will float or sink, or whether there is not enough information to make an accurate prediction. (In many cases, the last option represents the most scientifically defensible choice). Students must also provide the reasoning that underlies their answer. This task requires them to organize a system of judgments and conclusions somewhat along the following lines: *The larger cube has sunk because it is denser than the liquid in which it was immersed. The smaller cube, weighing the same as the larger, must be even denser and so will surely also sink in the same liquid.*

One learning goal that can be assessed through observing student performance on the above activity is the ability to conceptualize the density of solid objects. Density is a central scientific concept that underlies competent performance in all of the natural sciences that are to be studied in high school, and so we consider this a worthy learning goal. Students' predictions and reasoning concerning the density of the cubes can be characterized in part by using the following scale of competence:

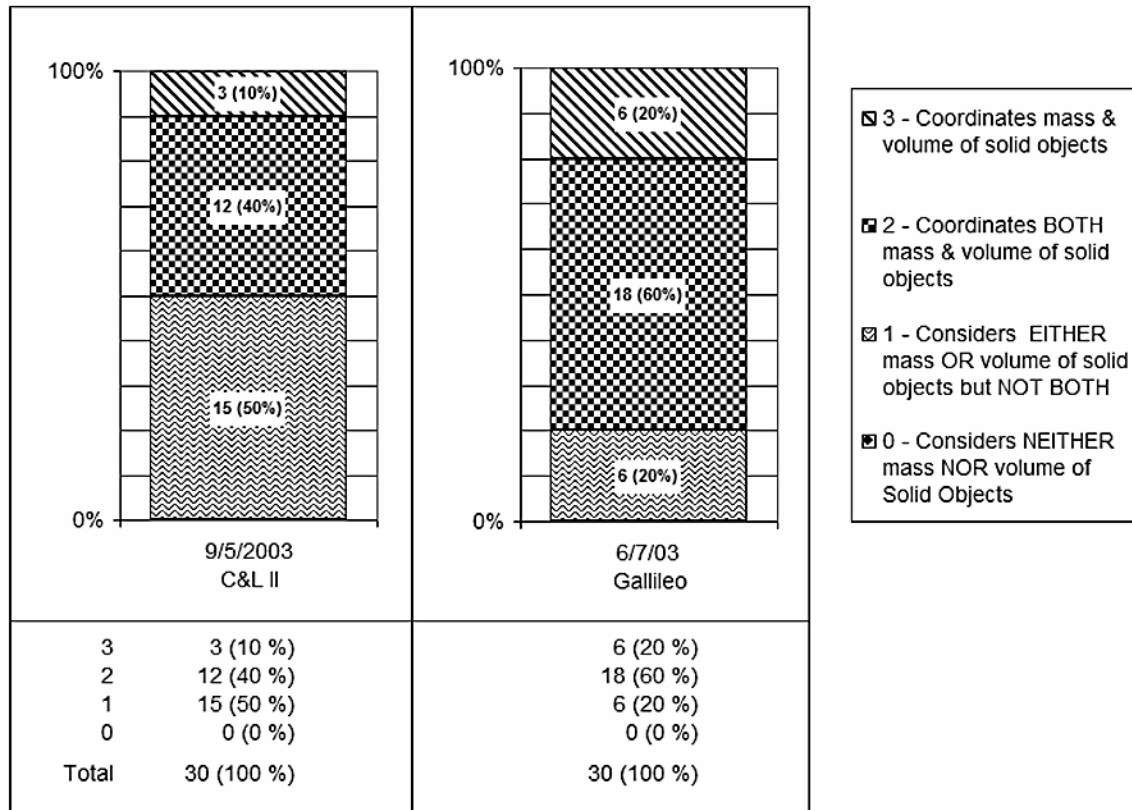
Conceptualizes Density of Solid Objects

- 3 Coordinates mass and volume of solid object
- 2 Considers both mass and volume of solid object
- 1 Considers mass or volume of solid object, but not both
- 0 Neither mass nor volume of solid object are considered

(A similar learning goal exists that deals with conceptualization of the density of liquids, and another that concerns understanding of the relative roles of the densities of the solid object and the liquid in floating or sinking.)

Students' performance on the task described above provides a basis for using the scale of competence to infer the students' conceptualization of density. The scale provides considerably more than a way of indicating that some students are doing better than others. It serves the diagnostic role of telling us, specifically, what students have attained or what they are lacking in knowledge or skill. It provides information that is immediately translatable into instructional action. From this information a teacher can form a picture of the ability of the class or of individuals to conceptualize the phenomenon. The figure, below, represents one possible visualization of the performance of a full class on the conceptualization of density of solid objects:

Conceptualizes Density of Solid Objects



This report provides graphic and numeric representations of the proportion of students who are functioning at each level of competence on the conceptualization of solid density scale at two points in time. This provides the opportunity to observe a progression in attainment and to consider whether expectations for student attainment were met or not. In the on-line information system from which this screen image was taken, one can ‘drill down’ into each of the scale levels in the bar graph to identify the individual students performing at each level, and thereby see who is most in need of additional instruction.

Thus, educational assessment can give the teacher information needed to plan an appropriate lesson directed to both the class and to individual students. The teacher has an empirical basis for lesson planning, but is, at the same time free to use discretion in creating an imaginative response to the facts of student performance. We recommend to teachers that such assessments be administered as a pre-test (to determine what needs to be taught and to whom), and over the course of instruction to gauge its success. A post-test can provide an ultimate evaluation of the degree to which instruction was successful.

We discourage teachers from grading student work based on attainment of learning goals such as this one. We recommend that teachers make a point of telling students that they will not be graded on the basis of whether their answers are correct or incorrect. If the teacher *must* assign a grade based on these assessment tasks, we encourage the teacher to assign the grade based on whether students complete all the items in the assessment, whether they provide neat and clear responses, and whether they use complete sentences (or try to) when they are asked to do so. There are many reasons for this; let us consider only three:

- 1) In order to have the best information for instructional planning we want to know what students actually believe and think, not what they think is the 'right answer' that will please the teacher.
- 2) If students are concerned about whether they have the 'right answer' for the teacher they will not be able to focus their full attention on the phenomenon and the problem at hand.
- 3) By not grading students for their attainment we protect the students' relationships to the natural phenomenon from contamination by the desire for social approval.

Thus, we have a truly educational assessment. By giving teachers careful preparation in administering such tasks and in objectively interpreting student performance, we achieve the rigor and reliability of information that is the hallmark of standardized tests. Furthermore, we do so without inducing many of the undesirable consequences associated with standardized tests. Teachers in public and independent schools are currently using these methods and finding them both illuminating and helpful.

Students cannot prepare for such an assessment task simply by memorizing concepts. Rather, they must be fluid and flexible in applying their concepts and reasoning capabilities to the phenomenon of interest. We believe that the development of the capabilities assessed through such activities requires that the teacher inspire interest, attentiveness, and personal engagement with phenomena. Those who may have 'crammed' the facts related to density and floating or sinking can find themselves worse off in dealing with such tasks than those who have not. Zachos et al (Zachos et al., 2000) found that students who were not explicitly working with a prior concept of density, but who were thinking in a lively fashion, were able to recognize the necessity for coordinating weight and size of objects, and so began to formulate a concept of density of their own, in order to solve the problem at hand.

We find that most secondary school students, at first, do poorly on the challenges associated with this assessment activity, even in the upper grades of high school. Although most have been taught as early as the eighth grade that floating and sinking depend on density, and most can even rattle off a formula for density [Density = Mass/Volume] and use it to calculate values, they do not effectively *apply* the concept of density to the phenomenon. Some do not take the mass or volume of the cubes into consideration at all, and claim that there is not enough information to know whether the small cube will float or sink because they don't know what the cube is made of. Others spin hypotheses concerning the properties and behavior of molecules in the cubes, but cannot relate them practically to the forces operating in the phenomenon. In neither of

these cases are the students able to apply what they have learned in school to the real life task. A test that rated students on whether or not they can recall or calculate the formula for density given above, or one that looks for a particle theory of matter as a higher level of explanation, in and of itself, would fail to show that the student's knowledge is very shallow indeed and, in fact, useless in the practical situation.

The Quandaries of Educational Assessment

We often hear that that some learning goals cannot be assessed. By contrast, I suggest it is more productive to think in terms of some learning goals being more difficult to assess than others. There are a number of reasons for this. First, we must consider that learning takes place in the inner life of the human being. Learning processes are by their very nature not observable, and that is true to varying degrees for learning outcomes as well. Consequently, in order to see whether goals for learning have been achieved we must often engage in a process of inference from the evidence of observable student performance, much as a scientist infers invisible laws and principles from observable natural phenomena. Assessment information may be derived from homework, lab reports, tests, essays, and even classroom conversations. But, like scientific investigation, sound judgments concerning student performance require careful observation, precise conceptualization, organization of observations, rigorous inference, and clear communication. Teachers are typically not trained to engage in these types of activities regarding student performance. It is possible to obtain honest and illuminating information about any learning goal, although in general we find that the less observable the outcomes of a learning goal, the more challenging will be the process of making valid inferences about the attainment of that goal.

We also hear arguments that we must wait some number of years – I've heard as many as 30 – before we can adequately assess the impact of our teaching. Perhaps this is true, yet there is little point in waiting until next year or after high school is completed to assess whether a child has attained the core concept of density. How the child performed in classroom discussions yesterday, what she wrote on her essay last week, and the question she is asking right now are the living basis for forming pictures of what she needs. These pictures can be the most crucial sources to inform decisions for planning and improving instruction. Eugene Schwartz (Schwartz, 1992) points out that, "if we accept the premise that the child is a being who unfolds his capacities over the course of time, then it follows that the most valid assessments to be made of a child's development are compiled over the long periods of time"(p.33). This, to my mind, is a good reason to assess over time, as often as is practical and wholesome, and in diverse ways as a basis for making valid judgments of student attainment.

Typically, the teacher does not have the time or the expertise to construct rigorous assessments such as the one described in the last section, and for this reason our association has made a commitment to develop assessments for core concepts and capabilities, assessments that are ready – with appropriate professional preparation – to be put to use in educational settings. There is an art and a science to preparing useful

educational goals, creating valid assessment, and conducting sound evaluations of educational activities. For example, it is important that educational assessment information be timely, reliable, and communicable in order for them to be useful to the teacher and the educational community.

In summary then, *educational assessment* is best seen strictly as the process of obtaining information that serves as evidence that *learning goals* have been attained. This precision of definition allows us to see clearly into the nature of the problems of testing and find ways to overcome them.

The Potential Contribution of Waldorf Education

There are a number of contributions that Waldorf Education can make to the educational community at large. These contributions include ways for teachers to develop a deeper understanding of the child, and a sense of the child's relationship to the larger scheme of things - to nature, history, and to the fulfillment of personal destiny in social life. Notions of the integration of thinking, feeling, and willing, sensitivity to cycles of wakefulness and sleep, and understanding of the inner dynamics of the human body in its threefold aspect, (Steiner, 1996) are all virtually absent outside of Waldorf Education. Concerns for the profounder aspects of human values and meaning take us beyond finite *learning goals*. They are in fact in a realm that can be considered the source and justification, the higher purpose of learning goals. Learning goals by their nature spring from human values and are tools for striving towards valued ends. Because of its explicit confrontation with such larger concerns Waldorf Educators holds the potential of offering a richer set of goals or expectations for learning and human development than is found in conventional educational settings. The step must be taken, however from higher values to setting and working consciously with finite learning goals (e.g. conceptualization of density). That a learning goal is finite does not mean that it is not important. Educational programs must always deal consciously and effectively with finite learning goals.

In the domain of pedagogy, Waldorf schools have pioneered the practice of child study, an approach that the larger educational community is just beginning to explore systematically. The depth of understanding of the stages of the child's development that characterizes Waldorf Education is also missing from most conventional classrooms. The argument for sensitivity to developmental stages in the educational life of the children was made cogently by (Inhelder, 1958) who provided a wealth of supporting empirical evidence for its importance. However, if anything, there seems to be increasing insensitivity to the question of developmental appropriateness of instruction outside of Waldorf and Montessori schools. Interest in promoting emotional intelligence, and multiple intelligences are two other areas that could be enriched with the fruits of almost a century of pedagogical practices pioneered in Waldorf schools. Attention to student temperament in educational diagnosis and planning are still to be fully discovered by the educational mainstream. Clearly Waldorf education has received and put into practice powerful indications for pedagogy. But pedagogic approaches and technique exist because of their anticipated contribution to the achievement of learning goals. In order to

establish whether these approaches and techniques are worthy and effective we must establish their contribution to the attainment of worthy learning goals.

What I find to be largely missing in Waldorf education is sound, rigorous knowledge – even consciousness – of what the child has and has not attained, that is, the extent to which goals for learning have been achieved. This is not surprising for a community which has, for the most part, not been working explicitly with educational goal setting and educational assessment as I have defined them. Waldorf educators should not avoid such educational considerations, and indeed attention to them would put Waldorf Education in a position of leadership and give it the possibility of bringing new insights and perspectives into learning and the purposes of education to the educational mainstream.

The Law of Educational Program Transformation

There is a law of educational program transformation and Waldorf schools are becoming increasingly subject to this law. The law states that: *The nature of the assessment that is used to evaluate an educational program determines the curriculum and instruction of that program.* This law dictates that when you use a certain test to decide whether an educational program is successful, teachers will start changing the program so that students will do well on the test. This law is the basis of the phenomenon of ‘teaching to the test.’ The fact that learning goals are not stated and worked with explicitly does not mean that they do not exist. What typically happens when this is the case is that the capabilities needed to do well on a test become the default learning goals. They then drive the instructional program in an unconscious and typically unintentional way. Students’ attention and ingenuity turn away from confrontation with the subject matter and the real world to “psyching out” the teacher and the test. In this way conventional tests blindly drive educational programs through their implicit learning goals.

To assure that Waldorf programs are effective mediums for the attainment of worthy learning goals – and to make possible the larger contribution of Waldorf Education to the renewal of the educational enterprise world wide, I recommend that a number of steps be taken:

1. Develop clear statements of the goals of Waldorf Education;
2. Establish by argument and otherwise, the worthiness of these goals; and,
3. Develop productive ways to assess the extent to which these goals are being achieved via *educational assessments*. The results of these assessments should be such that they would be of direct use to teachers in planning and improving instruction. They should also be such that they provide rigorous ways for teachers to communicate with students, colleagues, and their communities about student progress in attaining valued attributes, and how to best realize the students’ potential.

This would be an ambitious enterprise, but I believe, an enlightening experience as well. It might begin by considering core learning goals for year or semester long track classes.

In these classes, capabilities are developed that we assume students should have available to them for their work in the main lesson blocks. Information on the extent of achievement of these goals could be the basis of fruitful collaboration between track-class and main lesson teachers and a basis for rigorous fact-based child study leading to remediation or enrichment where needed. Because little has been done in this arena, the possibilities and freedom for creative approaches to both are quite broad. Another possibility would be to make explicit the learning goals implied in an existing curriculum such as *form drawing* that is unique Waldorf Education. The study of students' form drawings would have an added value in that form drawing provides information that can be used diagnostically to reveal student readiness for instruction in other curricula. In any case, it should not end up that the success of Waldorf education is judged by conventional standardized tests. Meaningful educational assessment could be Waldorf Education's response to the challenges and problems of standardization.

Conclusion

There is nothing that human beings can create that cannot be misused. So it is with methods of assessment and evaluation. Affecting change in human beings is always, as they say, 'a risky business'. The physician is enjoined to 'do no harm,' and we in education must take heed to follow the same prescription. By making sure that our assessments are purely 'educational' we can go a long way toward realizing the ideal of doing no harm and hopefully attaining some good as well. Ridding Waldorf schools of educational assessment would undermine sound practice in Waldorf education. It would impede the impulse of Waldorf education that is directed to deep knowledge of the child.

I have presented some ways to think about educational programs and particularly ways to think clearly about student assessment. I have argued that *educational assessment* based on *learning goals* is both necessary and beneficial. It is my hope and intention in presenting these ideas to begin conversations on productive approaches to educational assessment in Waldorf schooling. I look forward to being part of such conversations.

References

- Inhelder, B. a. P., J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.
- Johnson, M. (1977). *Intentionality in Education*. Albany, NY: Center for Curriculum Research and Services, State University of New York at Albany.
- Kohn, A. (2000). *The Case Against Standardized Testing*. Portsmouth, NH: Heinemann.
- Sacks, P. (1999). *Standardized Minds: The high price of America's testing culture and what we can do to change it*. Cambridge, Massachusetts: Perseus Publishing.
- Schwartz, E. (1992). Holistic Assessment In The Waldorf School. *Holistic Education Review*, 5(4), 31-37.
- Steiner, R. (1996). *The Foundations of Human Experience: compiled lectures from Steiner's life 1917-1919*. United States of America: Anthroposophic Press.
- Zachos, P. (2004). Pendulum Phenomena and the Assessment of Scientific Inquiry Capabilities. *Science and Education*, 13(THE PENDULUM: Scientific, Historical, Philosophical & Educational Perspectives), 743-756.

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. *The Journal of Research in Science Teaching*, 37(9), 938-962.