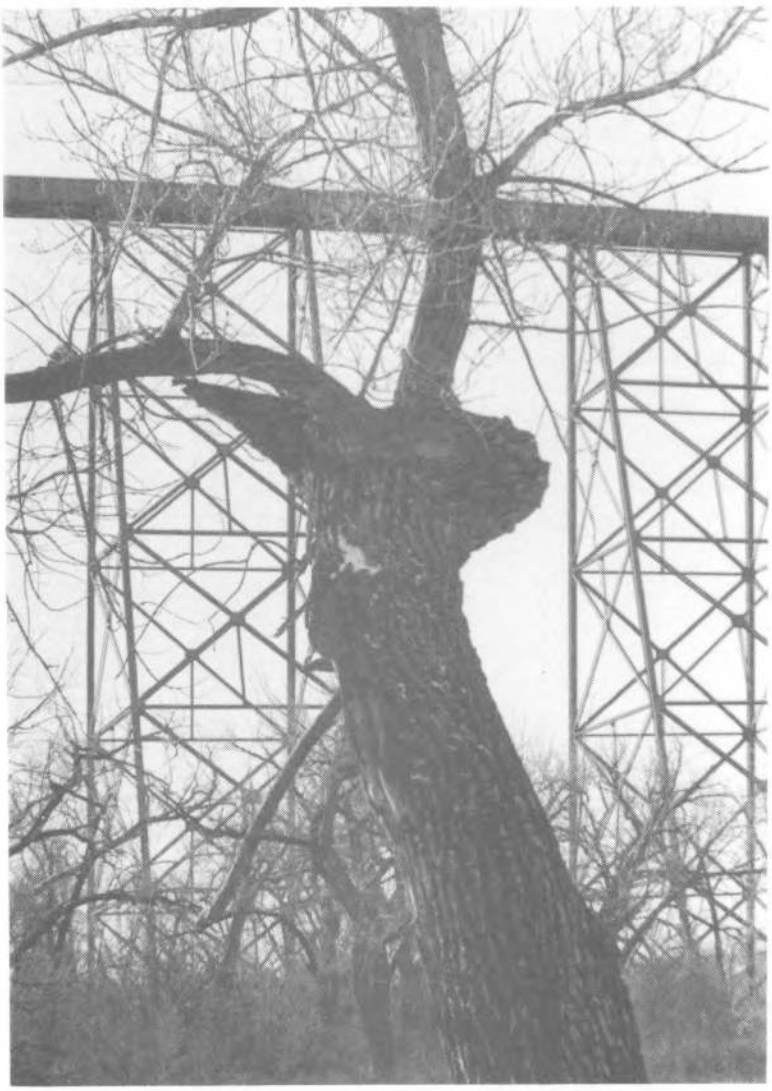

**Dancing Tree**

**Leslie D. McLean**
*Ontario Institute for Studies in Education*

# Achievement Measures Made Relevant to Pedagogy

**Abstract**

*Conventional achievement tests lack relevance to teaching and learning because they are based on inappropriate subject matter theories and the concept of achievement as trait. Functional language theories provide a base for relevant performance measures, but they also reveal that achievement is context and task dependent. Such dependence makes valid generalizations rare and application of item response models inappropriate. A functional approach here is outlined for mathematics and science in the middle school, where much improvement is needed. Content is carried by realistic problem sets, and achievement is defined as performance, documented in portfolios similar to the writing folders used in language classes.*

**Pedagogy** is defined in the *Compact Oxford English Dictionary* as, "the art or science of teaching." "The art **and** science of teaching" might be more accurate, or "the art and craft of teaching," but that is another story (McLean, 1985, for example). The purpose of this article is to explore what it means for achievement measures to be relevant to teaching and to suggest how the link can be made stronger. First, achievement measurement will be reviewed and analyzed to sketch out the problem. Next some encouraging progress in making language testing relevant to teaching will be described, which will lead to a final section suggesting how achievement measures can be made more relevant in other subjects, especially science and mathematics.

## Limitations to Prevalent Achievement Measures

Achievement is measured at three main levels of educational systems – the national level (provinces and territories in Canada), the local authority

(board or district), and the classroom level. At the national and local authority level, monitoring standards and accountability are the primary motivations, though the improvement of instruction is often cited. Some provinces, Alberta and British Columbia, for example, do construct achievement measures related to their own curricula, but so far there is no confirmation that teachers find the results directly applicable to their own teaching. Teachers have to make a long leap of inference from scores on common tests to the challenges of their own classroom, and the authorities who mandate the testing consider it sufficient if broad problem areas are identified for the professionals to solve.

In short, it is believed that monitoring and accountability do not require strong pedagogical relevance, a position that can only be defended if the measures used for monitoring tap outcomes toward which everyone is (or should be) striving. This defense is difficult if the measures are published tests that overlap provincial curriculum guides only a little, or one-dimensional scales that reflect outmoded concepts of the subject. More promising are province-wide surveys employing large item pools with good coverage of the provincial curriculum and contextual information from surveys of teachers, schools, and students.[1] Such surveys are not common, but the Ontario Ministry of Education has begun to carry out provincial reviews on a five-year cycle, using that methodology. Boards of education are invited to extend the provincial survey to their own jurisdiction by payment of a small fee per school. When extended to the school level (or better, the classroom level), such surveys can have considerable pedagogical relevance. The British Columbia learning assessments are models of such surveys.

To no one's surprise, measurement moves closer to pedagogy as it moves closer to the classroom, but classroom measurement that relies on test scores is still disconnected from teaching. There are two faces to the problem, neither of them discussed in the measurement literature. The first is that tests are based on inappropriate, usually unstated, theories of the subject matter. It may be unfair to blame the tests, because the concept of a theory of subject matter is new to most curriculum developers. Language theories and their application will be discussed below in the section on language teaching and testing. Some preliminary thoughts on theories of mathematics and science content will follow the language section.

The second obstacle to pedagogically relevant measurement is a concept taken over from psychology – achievement as **trait**. Under the influence of psychological testing, particularly ability or "intelligence" testing, the idea has become entrenched that school achievement has to be captured in a number, a test score. This has led to an enormous literature under the heading of "test theory," not a theory at all but an elaboration of

families of mathematical models for the analysis of test scores. The families, including the fashionable "item response" models, all have one characteristic in common, i.e., they are content free. From the classical model of observed score equals true score plus error to the logistic item response models, the same form is applied to a score said to measure reading comprehension as to a score representing achievement in mathematics computation. Test scores are analyzed by specialists in psychometrics who have no need to know what the test developers have set out to measure.

In the absence of subject matter theories that provide clear definitions of achievement, it is not surprising that achievement is construed as a general human characteristic, a trait that is stable across people and settings. Pedagogy, however, takes place within specific settings, among people who have both shared and unique intentions. Curriculum guides, quite rightly it seems, leave many important decisions to the teachers, with the result that the content and concepts presented to students vary in important ways from classroom to classroom. At its best, this system permits the skills of the teacher to be matched with the interests of the pupils, still respecting overall content guidelines.

The system makes sense, because in spite of constant efforts on the part of confident zealots, no specific content has been identified as the **necessary** curriculum. Many are now equally confident that there is no **necessary** curriculum, and hence that the learning required for productive living in any society has to be defined at a more abstract level than the school's operational syllabus (course of study). After several years of work, an Ontario study, for example, has found that "negotiating skill" is a key outcome to be attained through the curricula of the elementary and secondary schools (Russell, 1987). In other words, the school's course of study can be chosen in a valid way from a wide selection of alternatives, much as is now done when schools interpret provincial curriculum guidelines.

Such diversity is incompatible with the concept of achievement as a stable, general trait. The weak predictions provided by trait measures testify to the low utility of such constructs for understanding behaviour. For some, the utility has slipped to zero (Olson, 1986).

> There used to be a thing like verbal ability that explained things. It's gone. At least it's gone for me. All cognitive science is based on the notion that you have set procedures that you use for dealing with domains. Some of those procedures are applicable across domains and so on. And the task for psychologists, as I think for educators and others, is to find out just what that competence is made up of and

> what are the conditions under which you can help people sort out the major dimensions or considerations in that form of competence. There is still talk about spatial ability and verbal ability and so on, but that's a level of description that has very little explanatory value. Explanatory value comes from actually figuring out how they solve this task, or how they sort out what next to put down on their text if they're writing something. (p. 177)

This "explanatory value" is just what teachers seek to help them improve their teaching, what they get from questioning students in class and from homework exercises, but do not get from set tests. Diverse tasks, **test items**, are grouped by content, sometimes thoughtfully, but the outcome on which judgments are to be based is almost always a single number or score for each person. The thoughtful diversity built into the tasks disappears into the "sacred" test score. In spite of fifteen years of work with criterion-referenced testing, most educators still find meaning in test scores by comparing them with other test scores, in other words, with norms. Test scores tell us in general terms how learning is proceeding, but the combination of item results into a single number severs the link to pedagogy. It appears as if all of the pedagogically-relevant information available from tests is at the item level (Talesnick & McLean, 1987), but even at this level the tests fall short (Leinhardt & Putnam, 1987, p. 585). Measurement textbooks still emphasize the test score, however, giving primacy to summative judgment over formative. If mentioned at all, pedagogical relevance comes in as the unsubstantiated claim that test scores can be used to guide instruction.

It has been implied that inadequate theories of subject matter are one cause of the current predicament – theories of language, mathematics and science, for example – that not only suggest how the topic should be taught but also provide a definition of achievement and prescribe how this achievement should be measured. There is progress in the language field that may show us the way toward pedagogically relevant measurement in other subjects. That is the subject of the next section.

### Encouraging Developments in Language Teaching and Testing

Until the second half of this century, language teaching was based on a theory (not well articulated) that the structure of language was basic to its understanding and use. Teachers, therefore, tried to help pupils build up an understanding of structure by means of grammar, syntax, and spelling, and they tested discrete structural elements accordingly. About this time, interest in **foreign** languages increased dramatically as the global village became a

part of more people's reality. Reading other languages was no longer enough; people had to be able to speak and understand them and to communicate with them. In other words, the function of language began to take precedence over the structure. Writing emerged as a special pedagogical problem, combining the functional and structural elements in ways seemingly impossible to untangle. The pedagogy of writing remains a difficult challenge, mirrored by the difficulty of defining and assessing writing achievement (Oxenham, 1980).

It took several decades for functional approaches to emerge, and they are still emerging. In the late 70s, the new concepts converged in **communicative** language theories, containing direct implications for teaching and testing (Canale & Swain, 1980). The theories apply equally to mother tongue and to second, third, and other languages, since the emphasis is on the communication of meaning. Structure becomes the servant of meaning rather than the master. Once this shift is made, numerous implications follow, only a few of which can be mentioned:

- Context and purpose are recognized as vitally important. One communicates in a different way to different people in different places for different purposes. Teachers must therefore set up or take advantage of many different settings, showing and explaining why language varies with context and purpose.

- Communication is intimately related to knowledge and information, to the ideas and facts a person already has. Reading and listening are complex interactions between what is being read or heard and what one knows. Meaning does not exist in the abstract but instead is constructed in the process of reading or writing.

- The changes that go with learning are subtle, but they are continuous and cumulative. The pace is slow, happening over months rather than weeks or days, because of the many interactions in the process.

- Valid measurement must therefore involve similarly varied settings and purposes. It must take place over time periods of the order of months and years, and it must be sensitive to cumulative change. Achievement is multidimensional and is task and context dependent.

- Individual differences between students and between tasks will be great and important. Aggregate measures will have little or no pedagogical validity. Systematic, cumulative records of performance, such as the record provided by writing folders, are required.

Where these lessons are taken to heart, testing and assessment take a form which supports communicative teaching methods. The approach to language in the *Assessment of Performance Unit* in England, Wales, and Northern Ireland, for example, includes diverse materials, different writing modes, and linked speaking and writing tasks in several modes (Gorman, 1986; Thornton, 1986). In Ontario, the use of the writing folder has been mandated from grade one to the end of secondary school, and common examination procedures have been installed rather than common examinations at the end of secondary school. An international study of written composition is just being completed, following the collection of ten different samples of writing, and after subjecting them to various analytic scoring schemes. In 1984, the Department of Education and Science for England and Wales endorsed the use of *Records of Achievement*, comprehensive profiles of performance that pupils take with them from year to year and when they leave school (Broadfoot, 1986).

Not all jurisdictions have this view of language assessment. In the United States, the **National Assessment of Educational Progress** (NAEP) has issued a "reading report card," based on a single scale of reading proficiency that extends from age 9 to 17 (NAEP, 1985). The approach already has its critics (McLean & Goldstein, 1988), but it is exceedingly popular with decision-makers who demand simple answers. An American survey of reading performance in an out-of-school population used a broader sample of meaningful materials, but the results were then made to fit the same one-dimensional scale (Kirsch & Jungeblut, 1986).

### Implications for Science, Mathematics, and Beyond

When reading and hearing about developments in language theory and language teaching, one encounters a continuous interplay between structure and function. Function, however, is dominant, and structure is an essential support. People learn language naturally (from parents, relatives, others) with little or no recognition of the structures they are learning, but in order to generalize and to make learning more efficient, reference to structure is required. An important discovery was that teaching succeeds best by working from function to structure, and almost not at all in the other direction (Canale, 1983). Structure is now seen as the servant of meaning, not the master. Might this not apply to other subjects as well?

Both science and mathematics teaching are in somewhat of a turmoil. In Canada, two national studies have found only small amounts of science teaching and distressing levels of performance in the general school population (Science Council, 1984; Connelly *et al.*, 1985). Overall achievement in mathematics is not so distressing, but neither is it inspiring (McLean *et al.*, 1987), and the same problem areas exist now as have

existed as long as analyses of the curriculum have been made. The middle school (roughly ages 11-16) is the most problematic, with a great deal of review of arithmetic and little progress in it. Average student performance is poor on any but the simplest applied problems (McLean, 1982). The field would seem to be open for new approaches.

Mathematics and science curriculum guides are marvels of structure – abstract frameworks that organize the parts into a whole. The term **structure** is said to have been introduced to sociology by Herbert Spencer in the 18th century, and it now pervades social science. Structure serves as the unconscious infrastructure of phenomena. Structural linguistics dominated language analysis up to recent times, though it is notable that it was not prominent in semantics. Apparently meaning does not yield so readily to structural analysis. The concept made its way through Durkheim and Radcliffe-Brown to Talcott Parsons, whose functional-structural analysis saw function as a link between relatively stable structural categories. Structuralism has been seen by some as a spectre over social science, because of its ubiquitous use as a vague term (Assiter, 1984). It is similarly a spectre over curriculum where it blocks teachers' and students' view of the functional elements of their subject (as it did in language for many years). This is seen, for example, in the **applications** problems in most mathematics and science textbooks in the form of contrived problems with meaningless settings.

Attention will be restricted here to the middle school, specifically grades 7 and 8, since the problems are well recognized there. In the United States, **problem solving** was declared to be the emphasis for the 80s (National Council of Teachers of Mathematics, 1980), but evidence for implementation is lacking. Functionalists would concentrate on meaningful problems, giving attention to context and purpose, leaving until later the place of the problem in the overall structure of mathematics or science. Half of the students now fail to learn key elements of middle-school mathematics, so there is room for improvement (McKnight *et al.*, 1987; McLean *et al.*, 1987).

The following criteria for selecting school problems are an attempt to capture a functional approach to the teaching of mathematics and science:

1. The problem must have an interesting context that is either already known to most students or is easily recognized.

2. The mathematics or science needed to solve the problem is generally useful and central to the problem. The content is not a trivial or amusing sideline; it has an important function.

3. A satisfying solution is accessible to the majority of students to whom it is presented, but the problem has many more challenging dimensions for able pupils or for the later study.

4. Other versions of the problem are easily accessible for practice and for evaluation without boring repetition.

There is not time to explore the functional approach thoroughly, nor is it yet fully worked out. What is clear, however, is that achievement can be defined as performance on a series of problems that are meaningful to the students. Student evaluation is then properly based on a systematic, cumulative record of performance – a portfolio unique to each student, inspired by the writing folder. The pedagogy to which achievement measurement is to be relevant begins with the context and purpose, introduces the solution as an application of content, and proceeds to related problems in order to seek transfer and generalization. At least in the beginning, no attempt is made to teach problem solving explicitly, but a time is set aside for reflection on the solutions and approaches at the end of work on each problem set. More time is spent on fewer topics, a lesson suggested by several studies in recent years (McKnight *et al.*, 1987; Science Council, 1984). The approach suggests organization of cooperative work in whole class and small groups, an organization strongly supported in recent research (Johnson *et al.*, 1983).

## Summary and Conclusions

Missing or inadequate theories of subject matter and an inappropriate concept of achievement have been proposed to explain why prevalent achievement measures, mainly test scores, have little relevance to pedagogy. Functional language theories however, are showing great promise in guiding language pedagogy, and they also suggest related measures of achievement. The theories imply achievement as performance rather than achievement as trait, requiring recognition that achievement will be task and situation dependent. Such measures do not lend themselves to summary by test scores and, in particular, are inappropriate for fashionable techniques such as the application of item response models. Instead, the theories suggest the use of portfolios (reading, listening, speaking, writing, mathematics, and science) as systematic, cumulative records of performance.

Implications for other subjects were found in the concepts of function and structure that underlie communicative (functional) language theories. A functional approach to science and mathematics first sets problem clusters in a familiar context and motivates a solution that is accessible to the students. Further elaborations of the cluster give practice in the key concepts and skills, not striving for understanding of the overall

structures until after the middle school. Records of mathematics and science performance are linked directly to pedagogy. It remains to be seen whether a comprehensive pedagogy can be devised, but achievement in the middle school is sufficiently low that there is room for new initiatives.

An earlier version of this article was presented at Mt. St. Vincent University, at a joint seminar of the faculties of education of the College of Art, Dalhousie, St. Mary's, and Mt. St. Vincent Universities, Halifax, Nova Scotia, October 15, 1987.

## NOTE

1. An example of such a survey is the recently completed *Second International Mathematics Study* (Robitaille, O'Shea, & Dirks, 1982; McLean, Wolfe, & Wahlstrom, 1987).

## REFERENCES

Assiter, Alison. (1984). Althusser and structuralism. *British Journal of Sociology, 35*(2), 272-296

Broadfoot, Patricia. (1986). Profiling and the affective curriculum. *Journal of Curriculum Studies, 19*(1), 25-34.

Canale, Michael. (1983). From communicative competence to communicative language pedagogy. In J. Richards & L. Schmidt (Eds.), *Language and communication.* London: Longman.

Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1-47.

Connelly, F. Michael; Crocker, Robert K.; & Kass, Heidi (Eds.). (1985). *Science education in Canada. Vol. I: Policies, practices, & perceptions.* Toronto: OISE.

Gorman, Tom. (1986). *The framework for the assessment of language.* Windsor, Berkshire: NFER-NELSON.

Johnson, David W.; Johnson, Roger T.; & Maruyama, Geoffrey. (1983). Interdependence and interpersonal attraction among heterogeneous and homogeneous individuals: A theoretical formulation and a meta-analysis of the research. *Review of Educational Research, 53*, 5-54.

Kirsch, Irwin S. & Jungeblut, Ann. (1986). *Literacy: Profiles of America's young adults.* Reading achievement of household survey of 25-year-olds. Princeton, NJ: Educational Testing Service.

Leinhardt, Gaea & Putnam, Ralph T. (1987). The skill of learning from classroom lessons. *American Educational Research Journal, 24*(4), 557-587.

McKnight, Curtis C.; Crosswhite, F. Joe; Dossey, John A.; Kifer, Edward; Swafford, Jane O.; Travers, Kenneth J.; & Cooney, Thomas J. (1987,

January). *The underachieving curriculum: Assessing U.S. school mathematics from an international perspective.* Champaign, IL: Stipes Publishing Company.

McLean, Leslie D. (1982). *Report of the 1981 field trials in English and mathematics – intermediate division.* Toronto, Ontario: The Minister of Education, Ontario.

McLean, Leslie D. (1985). *The craft of student evaluation in Canada.* Toronto, Ontario: Canadian Education Association.

McLean, Leslie D. & Goldstein, Harvey. (1988, January). The U.S. national assessments in reading: Reading too much into the findings. *Phi Delta Kappan, 69*(5), 369-372.

McLean, Leslie D.; Wolfe, Richard; & Wahlstrom, Merlin. (1987). *Learning about teaching from comparative studies: Ontario mathematics in international perspective.* Toronto: Queen's Printer for Ontario.

National Assessment of Educational Progress (NAEP). (1985). *The reading report card: Progress toward excellence in our schools; Trends in reading over four national assessments, 1971-1984.* Princeton, NJ: National Assessment of Educational Progress and Educational Testing Service.

National Council of Teachers of Mathematics (NCTM). (1980). *An agenda for action – recommendations for school mathematics of the 1980s.* Reston, Virginia: NCTM.

Olson, David R. (1986). Mining the human sciences. *Interchange, 17*(2), 159-177. (With discussion.)

Oxenham, John. (1980). *Literacy: Writing, reading, and social organisation.* London: Routledge and Kegan Paul.

Robitaille, David F.; O'Shea, Thomas J.; & Dirks, Michael K. (1982). *The second international mathematics study: The teaching and learning of mathematics in British Columbia.* Victoria, BC: Ministry of Education, Learning and Assessment Branch.

Russell, H.H. (1987). *Final report: Survey of skills needed in the workplace* (Tech. Rep.). Ontario Institute for Studies in Education.

Science Council of Canada. (1984). *Report 36. Science for every student: Educating Canadians for tomorrow's world.* Ottawa: Minister of Supply and Services, Canada.

Talesnick, Irwin & McLean, Leslie D. (1987). *Student achievement in Ontario grade 12 and grade 13 chemistry classes: Report of the 1983 field trial of the Chemistry OAIP.* Toronto, Ontario: Minister of Education, Ontario.

Thornton, Geoffrey. (1986). *APU language testing 1979-1983: An independent appraisal of the findings.* London: Department of Education and Science.