

E Q A O

AND THE MISMEASURE OF SCHOOLS

Testing at what cost
and to what end?

BY JIM NEILL

Debate continues to rage about the value of province-wide testing in public education. One article will never settle the question of “to test or not to test.” Instead, since there are three important and basic evaluation attributes for tests; validity, reliability and utility, I will consider the Education Quality and Accountability Office (EQAO) testing in Ontario schools in terms of those attributes to try to address the question, “Testing for what purpose, at what cost, and to what end?”

Construct validity seeks to confirm that the test logically, empirically, and

accurately measures what is being evaluated. EQAO testing is used to measure an individual student’s literacy or numeracy skills. In isolation, this may be a valid test for a valid purpose. However, the Ministry of Education makes school scores publicly available, and they are widely reported in the media. Organizations such as the Fraser Institute and C.D. Howe, purporting to evaluate the performances of schools and boards of education, use the results to rank order schools. As a result, test validity and reliability are jeopardized by the sense of competition and the pressures to succeed that result from these rankings and

ILLUSTRATION: SOPHIE CASSON



the consequent test practices.

The first rule used to be, “Never teach to the test.” Tests should be an accurate sample of a set of skills. The skills should be taught, not the test. However, the EQAO tests have taken on such perverse status that schools that have core text shortages, where students or classes must share textbooks, spend already scant resources on drill workbooks which replicate the test. School libraries, which research has shown are a key to literacy, are understaffed and under-resourced. Class time is borrowed from Grade 9 and 10 courses to teach specific test writing skills. Curriculum is modified in order to more directly teach the specific types of questions found on the test. Even EQAO endorses and encourages this practice by putting two practice versions of the test online with sample questions from previous tests. Scores have steadily improved. But at least some of the improved scores may be an artifact of students now being taught more explicitly how to write the test.

By the same token, when the EQAO scores are made public, it has become a colossal “What did you get?” game for schools and administrators. Principals are seldom judged directly, however, the way in which parents judge schools reflects on them. How the community values the school similarly reflects on them. The kind of positive learning environment which is fostered and perceived is important. But no other single measure, with the possible exception of graduation rates, reflects so heavily upon administrators and schools as EQAO test scores and so the pressures on administrators to ensure their schools do well on these measures will continue.

The result in many schools is that the test takes on an enormous, looming presence. Too often the test score takes on such an overwhelming importance that within the rules, potential success is maximized. There are more scribes, CD readings and computer written tests each year. To qualify for such test modification, students must have an Individual Education Plan (IEP) and so more students are having an IEP completed in

Grades 9 and 10. Acknowledging differentiated learners and therefore encouraging differentiated instruction and testing is a positive thing. Many schools have 25 per cent or more modified EQAO tests written. Again, this improved and more individualized test administration may account in some part for the recent provincial improvements in test scores.

Some schools have made the test a monumental event. In order to have sufficient scribes and test administrators, other classes are cancelled. Senior students are given supervised or unsupervised independent study, educational films are viewed and physical education activities are done. One thing that meas-



urement and evaluation experts agree upon is that tests should be as unobtrusive as possible. The emphasis we put on the test, the way it dictates the school day, and the way we administer it, may well increase the degree of student pressure and add to the obtrusive nature already too inherent in these tests. The days of the test in many schools are definitely like no other on the school calendar.

The consequence of pressures to do well on the test has led to several unfortunate true and troubling anecdotes. One school that succeeded in raising scores dramatically was looked at more closely in order to identify best practices. It turned out that a superintendent, a principal, a vice principal, and a teacher had not followed instructions and had improved test results after the fact. All were punished with College of Teachers suspensions and two retired as a result of those suspensions. There is no justification for such behaviour, but it is testimo-

ny to the pressures that many administrators and educators feel to improve test scores and therefore enhance their school's public ranking.

At another school, a student died tragically and suddenly just days before the test. The principal, expressing compassion, allowed students who were traumatized and grieving to choose a deferral. However, students who do not complete the test with their cohort are deemed to have failed, and that resulted in the school being identified as one of the poorest in the region by a local newspaper, and one of the poorest in the province by the foundation that does such extrapolated rankings. No caring educator would have made such a public pronouncement. The public release of school scores by the ministry and the subsequent ranking of schools allows such a cold calculation.

All tests have a standard error of measurement. Basically this means that the test score is not a hard or exact number. So, a standard error of ± 3 means a score of 77 actually represents more accurately a score of 74 to 80. To draw conclusions and make rankings from scores within this range, or marks that have improved or fallen within this range from one test year to the next, is statistically invalid. With published rankings however, such judgments are being made routinely on minor variances.

Statistical reliability is also effected by the size of the school. A smaller Grade 10 or Grade 9 cohort will have less reliable statistical data than a larger cohort since it might have, from year to year, wider variances in EQAO scores than a larger school. Again, this may mean scores improve or decline as a statistical variance rather than as a measure of achievement. Yet the public rankings continue.

In his book, *The Mismeasure of Man*, Stephen J. Gould humourously coined the phrase “physics envy” to describe practitioners of “soft” sciences who long for hard data to prove or disprove theories. Physics envy leads people to put far too much emphasis on numerical data, even when the numbers are not hard or

exact. Clearly a test score of 75 does not represent the same accuracy as a measurement of .75 milligrams. Educational psychometric scores are not exact measures, but rather educated approximations. To treat them otherwise, in the absence of any other supporting measures or observations is a mug's game. Yet that is how our schools are being judged.

Could also argued that intelligence is a theoretical construct and that to take such a construct and then to treat it as an exact reality is reification. Literacy and numeracy are exactly such abstract constructs, and so shouldn't be given the same concrete value as a chemical compound. Personally, I celebrate the fact that plumbers, mechanics, and electricians can read specs and manuals that would make me appear illiterate. Some of them, however, might well fail to get the required mark of 70 to pass this year's EQAO literacy test.

One handy guide to reliability is known as "Lyman's Five Dimensions of Reliability." These dimensions are interesting when applied to EQAO testing.

The first is "examinee incurred" and includes student motivation, stress, mental alertness, ability to follow instructions, stamina, work habits, etc. It's worth noting that most students have never written a three hour test prior to EQAO. They also may be doing so in an institutional atmosphere that may seem obsessed with the test.

The second dimension Lyman calls, "examiner/scorer influence." This influence is minimized when the test is purely a paper and pencil task. When scribes and additional time are allowed, the potential for examiner error is introduced and reliability is diminished. EQAO strictly prescribes that scribes need to be exact, but they do allow the suggestion of reviewing the response. However unintended or well intentioned, scribe cuing may introduce error.

"Test content" is another dimension and is simply the items chosen. These are tested by EQAO to make sure they are adequately discriminating. Indeed, each EQAO test includes items that are being field tested, which students complete. This is done for item analysis but not for stu-



Imagine...

Moving to a new country...

Having to re-establish yourself as a teacher...

Having to learn about a new education system...

Host a Newcomer!

Model the role of an Ontario teacher for an internationally educated teacher...

Provide an opportunity for an internationally educated teacher to observe an experienced Ontario teacher in action. There are no evaluation expectations of the classroom teacher. Simply welcome the internationally educated teacher for a two-day visit to your classroom.

For more information about this opportunity, contact the Project Manager:

Carol Norton-Sargent, 416-961-8800 ext. 420, 1-888-534-2222, ext. 420, cnortonsargent@teachinontario.ca

Coach a Newcomer!

Help open a classroom door for an internationally educated teacher...

Volunteer to provide guidance and support to a newcomer teacher for four hours a month over a period of six months through meetings, e-mail and phone calls. Share knowledge of the classroom dynamic and the role of the teacher in Ontario.

For more information about this coaching opportunity, contact:

In Ottawa: Haley Maunder, LASI World Skills, 613-233-0453, ext. 365, haley@ottawa-worldskills.org

In Toronto: Lourdes Zelaya, Skills for Change, 416-658-3101, ext. 239, zelaya@skillsforchange.org

In Windsor: Kelly Todd, Windsor Women Working With Immigrant Women, 519-973-5588, ext. 23, kelly@www.wiw.org



TEACH IN ONTARIO

Help for internationally educated teachers

Funded by the Government of Ontario and the Government of Canada through Citizenship and Immigration Canada



dent evaluation. This accounts for at least a portion of the three hours of the test.

The fourth dimension is “time influence.” This recognizes that improvements happen over time. So the second writing of the test the following year, as is the case in EQAO literacy, will clearly effect the results, usually positively.

“Situation induced” is the final dimension of reliability identified by Lyman. This has a stronger influence on less experienced or less motivated examinees. Environmental distractions such as storms, flickering lights, noise, etc. influence reliability. Again it’s worth noting that EQAO test takers indeed are inexperienced and the daily routine of their school may well be drastically altered on test day.

The original question was “Testing for what purpose, at what cost, and to what end?”


Individual EQAO scores are not shared in a timely way with educators. The test is primarily used to identify the student. The information shared is only remotely diagnostic. In other words, the information that is shared comes months

after the test and gives little insight into the specific skills that are deficient. So it names the unsuccessful student, but fails to give informed insight into instruction. There are far better diagnostic math, reading and writing tests that could help educators plan instruction. Such testing for teaching would be far more beneficial. EQAO fails as a test for teaching.

The provincial cost of EQAO testing is substantial, however, there are also hidden costs to schools. Human resources are expended in test preparation and administration. School curriculum is adjusted. Class time is spent in test preparation. Resources specific to testing are purchased. Yet, apart from naming students who fail, there is arguably little positive information that helps teachers improve instruction.

The most damning question, however, is “to what end.” Despite the questionable validity of EQAO in rating or ranking schools, this is unfortunately for what the test is best known. The unfortunate outcome of such public rankings is the “naming and shaming” of schools. Even CBC News, a trusted and reputable news

source, announced the newest Howe Institute rankings in a ten second sound bite that merely named the two lowest ranked schools provincially, with no context or critical analysis. This will be replicated across the province when local news outlets name schools each time the EQAO, Howe or Fraser Institute release their reports.

Clearly, we need to have an open and frank discussion of the merits of provincial testing in public education. Let’s ensure that any test is a valid and reliable measure of a student’s skills. These scores, however, have no place in the public ranking of schools. Tests should never be used to make extrapolated and statistically invalid claims. Doing so may benefit the agenda of the Howe or Fraser Institute (both of which favour public support of private education), however, it does little to improve student performance or to enhance public education. 

Jim Neill, a teacher librarian from OSSTF/FEESO District 26, Upper Canada, is a member of OSSTF/FEESO’s provincial Communications Political Action Committee.



Make Canada Your Classroom

With YMCA Youth Exchanges Canada, your school group can explore Canada while learning leadership, communication and teamwork skills. Students connect with new friends from another community and discover the history, geography and diversity of another part of the country.

10 to 30 participants between the ages of 12 and 17 are matched with a group from another part of Canada.

We're now accepting applications for the 2008/09 school year, so call 1-877-639-9987 or visit www.ymcaexchanges.org today for more info.

 Canadian Heritage Patrimoine canadien





