Queen's University

Education Letter

A publication of the Faculty of Education and the Education Alumni Committee Spring/Summer 2011

Testing and Ranking

MESSAGE FROM THE GUEST EDITOR

Don Klinger, *Faculty of Education, Queen's University*

Public education continues to reflect the shifting educational, political, and social perspectives across Canada. Even though public education is a provincial/territorial responsibility, and the individual perspectives of each province/territory result in differing educational systems, there is a surprising level of commonality across systems. As an example, every province and territory now administers a large-scale assessment (testing) program. While such assessment programs have not had a constant presence in Canada, they do have a long history. As public education became more commonplace during the 19th century, eventually becoming free and compulsory, there was a need to establish and maintain standards for student achievement. Educational reformists such as Ryerson promoted the widespread use of centralized examinations for the purposes of High School admission or certification. These examination programs became particularly important mechanisms of central control and authority, enabling the central authorities, primarily government and universities, to exert control over school curriculum. They also provided a measure of quality assurance.

It has now been 140 years since free and compulsory education was established in Ontario, and large-scale provincial testing is once again a part of public education. However, the roles and purposes of such testing in Ontario and the rest of Canada have changed and continue to change. Initially, these tests provided a broad measure of system monitoring in the elementary and early secondary grades, and in some provinces, a common measure of student achievement for students graduating from high schools. Currently, provincial

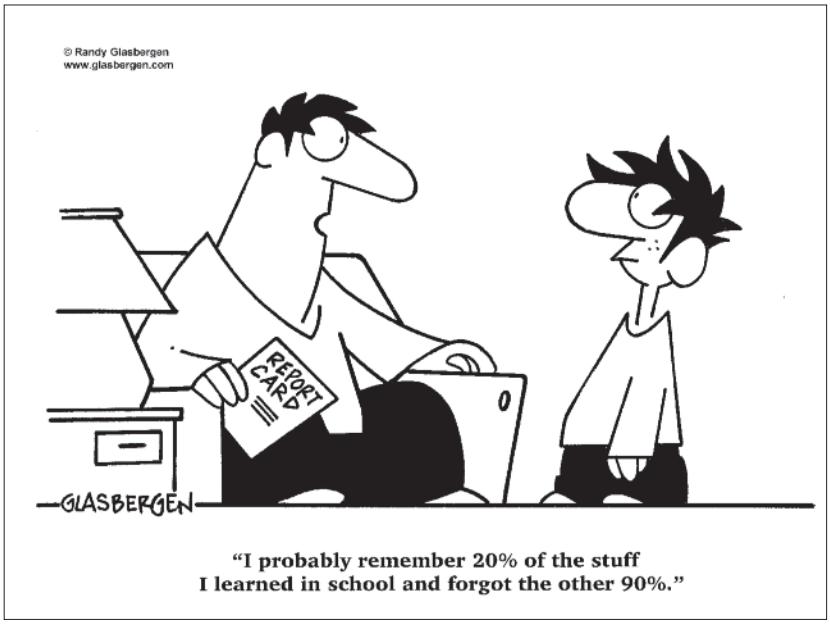
Três Semanas, by Maria Tomaselli Cirne Lima, 1993

... continues on page 2

… continued from page 1

and territorial testing programs commonly provide the data and information that form the basis of school accountability and improvement planning. In Ontario, these tests are also used to direct resources and support to schools having larger proportions of students who do not meet provincial standards. Along with these intentional uses of large-scale provincial testing programs, the publicly available school level results have also found other uses. This issue of the *Queen's Education Letter* deals with one such use, the publication and use of school rankings based on provincial testing results.

School rankings based on external examination results are not new. In England, these rankings are posted in the form of "League Tables." In the United States, rankings are used to identify poor performing schools, typically for the purposes of sanctions. School rankings in Canada began in the 1990s in Alberta and British Columbia. They were completed by external agencies, most typically the Fraser Institute, although it was not the first to do so. These rankings were based on the premise that schools differ in the quality of education they provide to our children and that rankings can be used to improve education. As David Johnson writes in his article, *Elementary School Assessments in Ontario are Great Value for Money*, such rankings have a potential benefit for parents and educators. These rankings, when done properly with respect to context, have the potential to identify promising educational practices while also assuring our public that education is providing value. He also notes that the cost of such testing programs is reasonable and certainly worth the investment given the amount of public money that is put towards education.



"I probably remember 20% of the stuff I learned in school and forgot the other 90%."

…continues on page 3

… continued from page 2

In contrast, W. Todd Rogers (*Should There Be School Rankings? No!*) argues that the contextual variables that Johnson and others use to more fairly compare schools are insufficient to provide fair rankings. While Rogers and Johnson both agree that such rankings result in winners and losers, they disagree on the consequences of such rankings. Johnson argues that "winners" are those schools that have higher levels of achievement in comparison to similar schools. The "losers" are then able to learn from these higher performing schools. Yet Rogers questions the negative impact of rankings on students, teachers, principals and parents that are in those schools that are "losers." Based on the *Principles for Fair Student Assessment* and the *Standards for Educational and Psychological Testing*, Rogers argues that in the absence of key questions that have yet to be answered, such rankings are irresponsible and likely unethical. Rogers also points out that Ministries of Education do not promote or condone the use of school rankings.

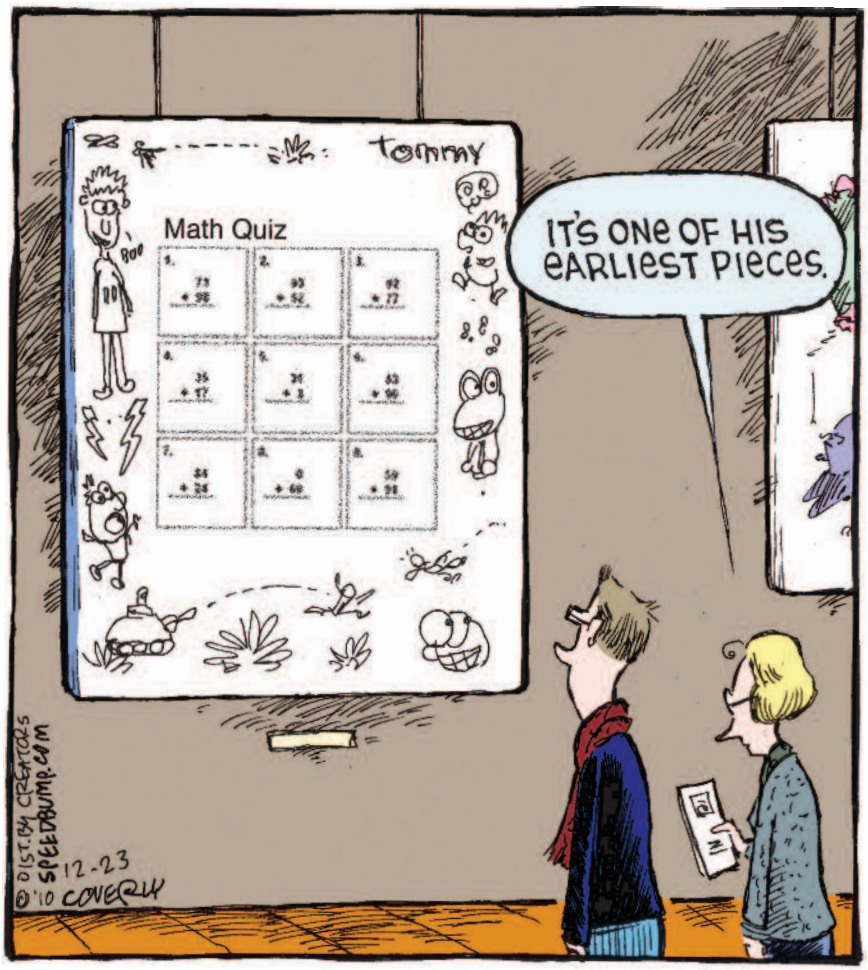
Anita Ram and Ben Levin (*Testing and School Ranking*) and Lynne Hollingshead and Ruth Childs (*The Perils of Reporting the Percentage of Students At or Above Level 3*) have approached the issues of school ranking from a different perspective. Together they build on the volatility in school scores described by Rogers, to identify the sources and issues associated with school score variability. While arguing in favour of the structure of large-scale testing programs that exists in Ontario, Ram and Levin caution against the adherence to annually published scores, scores that have been found to fluctuate considerably from year to year. This lack of stability makes it unreasonable to rely on single predictors as measures of school performance, let alone for the purposes of school ranking or the identification of high or low performing schools. Lastly, Hollingshead and Childs challenge the foundations on which the rankings in Ontario are based. Extending the issues of measurement error raised by Ram and Levin, Hollingshead and Childs provide an excellent example of how schools struggle to correctly interpret test results and ignore the inherent errors in the results they receive. They go on to illustrate the limitations of using the proportion of students who are meeting the provincial standard as the sole measure of good or improving instructional practices or student achievement.

The authors who have contributed to this issue of the *Queen's Education Letter* represent some of Canada's most important researchers in large-scale testing and educational policy. They are familiar with the Ontario education system and have been influential in shaping the current policies and practices that exist in today's schools. While they may differ in their perspectives regarding the use and impact of school rankings, they all believe that large-scale testing programs do have the potential to benefit our children and teachers. They all acknowledge the inherent challenges that exist in large-scale testing, and have provided their own recommendations to improve the manner in which results are determined and reported. Johnson may be the lone author who supports school ranking, but the conditions under which he supports these rankings are consistent with the issues the other authors identify.

… continues on page 4

… continued from page 3

Not surprisingly, the issue of public school rankings is controversial, especially when these rankings are based on external measures. These external measures provide valuable information to educators and the Ontario public, but the reporting and use of such information must first and foremost support teaching and learning. The thoughts of these authors demonstrate their efforts to ensure that such large-scale testing programs serve these two important functions.



Printed with the kind permission of Dave Coverly

RESOURCES

Hollingshead, L., & Childs, R. A. (2011). Percentage of students above a cut score: The effect of group size. *Educational Measurement: Issues and Practice, 30*(1), 36-43.

Johnson, D. (2005). *Signposts of Success*. Aurora ON: C. D. Howe Institute

Klinger, D. A., & Rogers, W. T. (2011). Teachers' perceptions of large-scale assessment programs within low-stakes accountability frameworks. *International Journal of Testing, 11*, 122-143.

Klinger, D. A., DeLuca, C., & Miller, T. (2008). The evolving culture of large-scale assessments in Canadian education. *Canadian Journal of Educational Administration and Policy*, *76*. **http://www.umanitoba.ca/publications/cjeap/articles/klinger.html**

Elementary School Assessments in Ontario are Great Value for Money

DAVID JOHNSON
Professor of Economics, Wilfrid Laurier University, Waterloo, Ontario

I am often asked why an economist looks at the results of EQAO elementary school assessments. My usual response is that education is the second largest economic activity in the country and whether resources are well spent within that sector is of central interest. Ontario spent 33 million dollars in 2010 on the Education Quality and Accountability Office (EQAO). I argue that even if every cent of this money was spent on the assessments carried out in Grade 3 (primary) and Grade 6 (junior), this would be money well spent. Since the EQAO also operates the literacy assessment in Grade 10, the mathematics assessment in Grade 9 and the national and international assessments, not all the budget is spent on the primary and junior assessments.

Thirty three million dollars sounds like a lot of money but it is not. The EQAO 2009-10 Annual Report (page 13) presents the number as $17 per student per year in a system that spends approximately $10,000 per student per year. Another way to present the number is to divide 33 million by the 4923 schools in 2007-08 (the latest year where complete numbers are posted) to obtain $6703 per school. There is absolutely no truth in the assertion that the EQAO program is a reduction of even one tenth of one teacher per school. Cancelling the EQAO program would not give every school a music teacher or a librarian. Further, if the universal assessment program were cancelled and replaced by a program where random groups of students were assessed in a given year, the usual alternative promoted by the teachers' unions, many or most of the EQAO costs would not go away. The assessments would still have to be created and administered, with the only real financial gains being in the reduced amount of marking. The low cost per assessment, either per student or per school is the first part of an argument that the elementary school assessments in Ontario are great value for money. The second part of the argument relates to the benefits. I find three large benefits to a publicly reported and available elementary assessment program administered to all students at the end of the primary and junior divisions.

The first and most important benefit is that the EQAO provincial assessments facilitate a conversation between parent and teacher where the parent is empowered. Teachers argue correctly that they write report cards on the students' capabilities based on much more than the provincial assessment results. Of course they do. And if the report card and the provincial assessment results match, then the conversation between teacher and parent is straightforward. However, it is worth pointing out that each individual teacher has an incentive to exaggerate a student's abilities. Why? It is less trouble for a teacher (and the Ministry of Education) to provide words on report cards that state all is well. Fewer parents will complain. Any potential problems can then

… continues on page 6

… continued from page 5

simply be passed on to the next teacher. Lest this be considered an anti-teacher remark, I note the same incentive occurs in our universities. Every individual instructor has an incentive to exaggerate grades. This is not a criticism of teachers or professors, simply an observation about human nature. You may have had a teacher or professor who systematically gives grades lower than other instructors; that person is simply "mean". My students occasionally tell me I am such a professor.

It is relatively easy to imagine a situation in which a parent does not agree with their child's classroom report card. The parent may feel the report card grade either exaggerates the child's abilities or understates the child's abilities. Without evidence from a large-scale external assessment any such conversation basically ends with the assertion by the teacher that they observe the child more completely than does the parent and the teacher's grades are correct and the parent is wrong. The teacher can make the same statements about the child to the principal. With the presence of EQAO assessment results, a principal can ask the classroom teacher(s) to explain the difference between the EQAO result and the classroom assessments, especially if such differences existed for most of the class. If the assessment results do not match the classroom reports for a large group of students at the school, then the principal can have an important conversation with the teacher or group of teachers within the division. If we are going to spend $10,000 a year on a student, surely it makes sense to spend $17 every 3 to 4 years as a quality check.

The second benefit of the EQAO assessment process is that you can provide valid school-level comparisons. My research (there is a partial list at the end of this article) suggests that variations amongst outcomes across elementary schools in Ontario are not simply the result of variation in the background of students at schools. I have done similar research for elementary schools in Alberta and British Columbia. The background of students at schools is measured by linking the postal code addresses of students at a school to the social and economic data collected in the long-form census. It is clearly the case that the percentage of students who achieve at or beyond the expected level of performance (Level 3) is often substantially different between two schools where students' backgrounds at both schools are similar. Good teaching and good management of teachers by principals and even good management of schools by boards can be identified using the data produced by EQAO assessment results.

I need to make a sharp distinction between my analysis, undertaken on behalf of the C.D. Howe Institute, and the Fraser Institute presentation of school results in the same provinces. One distinction I like to make is Fraser "rankings" versus C.D. Howe "ratings." Both presentations present some schools as "winners" and others as "losers", that is, some schools are systematically better than other schools. However the Fraser Institute's use of the word "rankings" implies that the first school is somehow much better than the tenth school. This is simply not true. I do not do this. In my system of "ratings," I try to make it very clear that the limitations of the data and, indeed the

… continues on page 7

… continued from page 6

limitations of any valid statistical methodology, is that you can identify very clearly some schools where a large group of students over a number of years either substantially under perform or over perform when compared to other schools that teach very similar students, that is, a group of students with the same social and economic background. Thus my analyses identify schools that can be used as model schools where best practices are exhibited. These are the winners. There are also schools where the analyses of the data indicate substantial improvement is clearly possible. These are the losers. But my analyses are fair in that the comparisons are made only between schools where the students share the same background. This is simply not true of the Fraser school rankings. The Fraser rankings are unfair and not very useful.

The third important benefit of the EQAO assessment process is that there is a measure of outcomes for elementary and secondary students that can be compared across schools. Without such a measure, researchers in education, economists as well as education faculty and researchers within school boards, have no outcome measure at the student level that is clearly comparable across schools. I have an ongoing project asking if students who pass through middle schools do better or worse on the Grade 9 mathematics assessment than students who remain in the same school to the end of Grade 8. This is an important question as the Ontario school system shrinks and schools must be closed. Other researchers use this data in other important and useful ways.

The EQAO assessment program is a low cost way to produce extremely valuable independent data on teaching outcomes at both the individual and school level in Ontario. It should continue. By any reasonable measure, the benefits of province-wide assessments vastly exceed the costs.

REFERENCES

Education Quality and Accountability Office, Driving Purposeful Improvement: 2009-2010 Annual Report, Toronto: Queen's Printer for Ontario, 2010

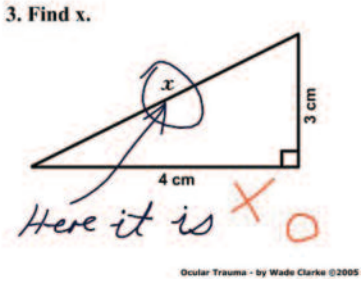
Johnson, David R. (2005) Signposts of Success: Interpreting Ontario's Elementary School Test Scores. C.D. Howe Policy Study 40. Ottawa: Renouf Publishing Company Limited.

Johnson. David R. (2005) An Update to Signposts of Success Ontario's Best Public Schools, 2005/06-2007/08, C.D. Howe E-Brief, August 18, 2009. **http://www.cdhowe.org/pdf/ebrief_85.pdf**

Johnson. David R. (2008) "Head of the Class: A Comparison of Ontario School Boards by Student Achievement" C.D. Howe Institute E-brief, August 27, 2008. **http://www.cdhowe.org/pdf/ebrief_39.pdf**

Should There Be School Rankings? No!

W. TODD ROGERS
University of Alberta



Printed with the kind permission of Wade Clarke

One of the more controversial issues in education today is the use of provincial achievement results to rank schools. There are those who say school rankings are justified, while others say they are not. First it should be recognized that what started out with rankings based simply on school achievement results has gradually been replaced with school rankings based on augmented school results where variables available from agencies like Statistics Canada and that are known to be related to student performance have been accounted for. However, the variables obtained are status variables and not variables that can be manipulated. Thus the augmentation is insufficient. Further, the method used by one agency is flawed due to the aggregation of data across grades and years. Inspection of the data reveals marked volatility of performance across subjects and years. But first, what is the stance of the agencies that develop, administer, score, analyze, and report assessment results regarding the ranking of schools?

Let's take Alberta, the province I am most familiar with, as an example. The following quotation, taken from the 2010-2011 General Bulletin (Alberta Education, 2010) provided to schools clearly lays out what the results of the provincial achievement testing program are to be used for:

> School and authority results from provincial tests are best interpreted within the context of local quantitative and qualitative information. …Personnel at the authority and school levels are in the best position to appropriately interpret, use, and communicate school authority and school results in the local context (p.4).

Nowhere in the General Bulletin is it stated that the results should be used to rank schools in the province. Further, this position is stated in the letter from the Minister of Education to the Superintendents of the jurisdictions within Alberta when the results are provided to the school jurisdictions and the schools within Alberta.

What the agencies that rank schools using data obtained from the provincial agencies fail to do is establish the consequential validity of the decision to rank schools and provide the rankings to the public. With school rankings there are winners and losers. Those who work in the field of education have pretty good knowledge about which schools will rank among the top ranked schools and which schools will rank among the bottom ranked schools in the province. But, what is the impact among the principals and teachers in the lower performing schools who daily work and provide learning opportunities taking into account the learning needs of the students in their schools?[1] What is the impact on parents and, perhaps, students in these schools who know that the students are receiving good instruction? These questions have not been systematically answered. In the absence of this data to answer these questions, the agencies that complete the analyses and the news outlets that provide the rankings

… continues on page 9

NOTES

1 The Alberta Teachers' Association estimates that approximately five teachers per year ask to be transferred from a grade that has testing to a grade that does not, citing the pressure they feel when the ranking results are printed in a newspaper or presented during television news cast. Based on this question, the ATA likely will be undertaking a more systematic study of teachers requesting a transfer and of principals asking their strongest teachers to teach classes who will be completing the provincial tests (personal communication G. Thomas, March 24, 2011).

… continued from page 8

to the public are violating all the standards (e.g., Part 4, *Principles for Fair Student Assessment Practices for Education in Canada*, 1993; *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement and Evaluation, 1999)) that have been established to ensure fair and equitable measurement and evaluation. In short, they are acting irresponsibly and likely unethically. Often these schools serve students who are not from the most advantaged homes. Many teachers, in light of the public rankings and the failure of senior educational officials to exercise leadership and have the rankings stopped, ask that they not be assigned to teach classes at a grade level with a provincial test.

What is known is that the provincial test results are discussed by teachers prior to the beginning of school. The teachers in each division (Primary, Intermediate, and Junior High) meet together for half a day to review the results for their school, the jurisdiction in which the school is located, and the province to see if and which changes need to be made to address the weaknesses identified. They complete this review taking into account the contextual and student factors for the school, and with the knowledge that the items included in the provincial tests are for curriculum objectives that can be assessed using paper and pencil tests and not all the curriculum objectives for which instruction is needed and that students learn. The factors considered are greater in number and wider in nature than the data used by the independent agencies who report school rankings augmented by easy to obtain, but limited in scope, variables.

The position taken in this short paper is that school rankings provide a disservice to schools and to principals, teachers, and students in the schools. This is not to be taken to mean that provincial achievement tests should be discontinued. As mentioned in the previous paragraph, results of provincial assessments are used by provincial officials in the area of curriculum as well as in the area of assessment, jurisdiction officials responsible for curriculum and instruction in the jurisdiction, and by school principals and teachers. Additionally, teachers in schools with only one class in a grade or only one section of a course with provincial assessments use the provincial and jurisdiction results as benchmarks to see how they are doing – these are the only benchmarks for these "single" teachers. Thus there are solid and relevant educational reasons for provincial assessments that are referenced to the provincial curriculum and that assess with relevant items that represent the measurable curriculum objectives. By doing so, and reporting the assessment results in a way that jurisdiction and school educators can easily interpret and use the results, a useful and meaningful service to foster student learning is being provided.

REFERENCES

Alberta Education (August, 2010). *General Information Bulletin, Achievement Testing Program, 2010-2011 School Year*. Edmonton, AB: Alberta Education.

American Education Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Education Research Association.

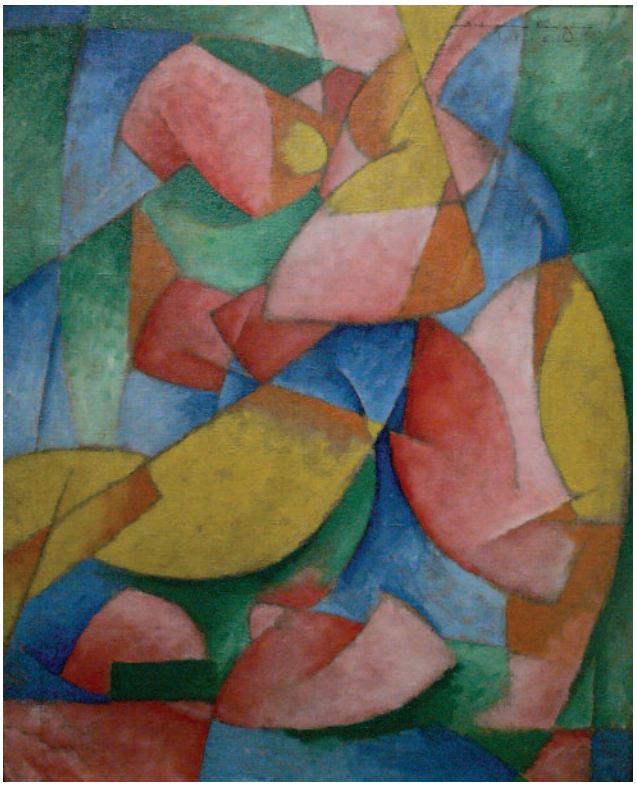
Principles for Fair Student Assessment Practices for Education in Canada. (1993). Edmonton, Alberta: Joint Advisory Committee. (Mailing Address: Joint Advisory Committee, Centre for Research in Applied Measurement and Evaluation, 6-110 Education Building North, University of Alberta, Edmonton, Alberta, T6G 2G5). **http://www2.education.ualberta.ca/educ/psych/crame/**

Testing and School Ranking

ANITA RAM and BEN LEVIN
Ontario Institute for Studies in Education, University of Toronto

Rating and ranking schools based on annual, external test scores is among the most controversial issues in education. Proponents argue that citizens, who fund the school system and entrust schools with their children, are entitled to know how well children and their schools are doing in meeting key academic goals. Opponents worry that the results have a substantial degree of error and that they may be used to exacerbate inequalities among schools. Both sides are right – and wrong. However the debate often yields more heat than light, with points being made that are not helpful to genuine public understanding.

Let's start with the argument for publication of school by school results. As public institutions, schools do owe some form of reporting on their achievement of the goals expected of them. Simply put, it is very difficult to generate improvement and to track progress if schools are unaware of their initial status. And while it is true that there are many important goals of schooling beyond literacy and numeracy, it would be difficult to argue that literacy and numeracy are *not* basic goals of schools; the argument for some public reporting of student outcomes in these areas is very strong.



Figure, 1913, oil on canvas by Morton Livingston Schamberg

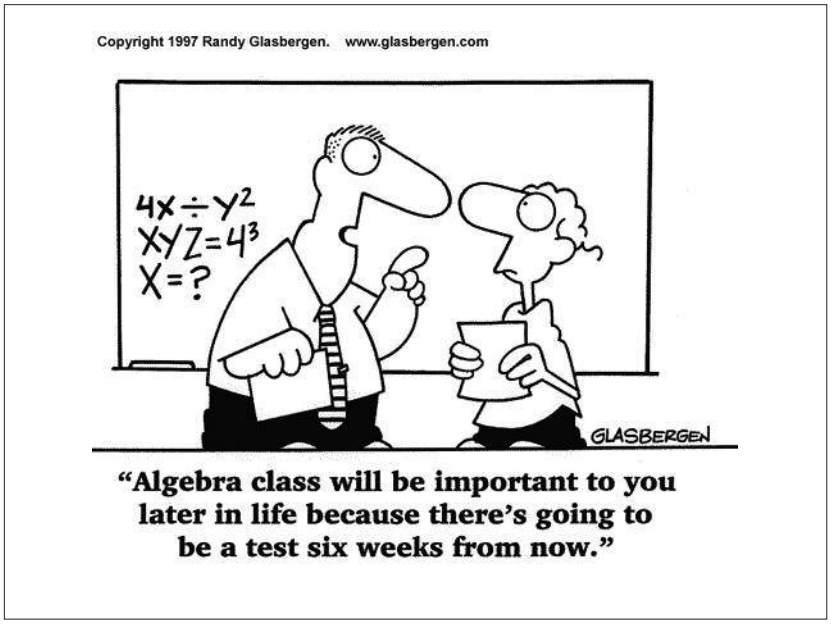
… continues on page 11

… continued from page 10

Some have proposed using samples of students from randomly selected schools instead of testing all students. However, sampling does not provide information on individual schools. The risk of eliminating or reducing these tests then is that the system would be unaware of the varying levels of performance within and between schools. There is a considerable risk that it will be poor and minority students who are disadvantaged by a lack of data on all schools and students.

Of course with testing there is a risk that schools will, due to the perceived pressures around test results, narrow their teaching. In the case of Ontario, such an approach by a school would be twice as bad a choice, because, aside from the flawed lessons it teaches students about the meaning of education, better EQAO performance requires a broad approach to the entire Ontario curriculum with a strong emphasis on higher order skills. These skills cannot be obtained through typical test preparation strategies.

It also seems ironic that schools, which are, after all, in the business of rating and ranking students, are so resistant to being rated themselves. Every day, schools assign grades to students and use the results to place students into classes or programs. These grades have the same, if not greater, problems of validity and reliability as do external test scores (it is well established that school marks are influenced by students' ethnicity, gender, and behaviour), and can have very powerful consequences for students. Yet there does not seem to be the same concern expressed about the potential unfairness of such classroom based grades.



… continues on page 12

… continued from page 11

But the critics of public reporting of external tests also make some strong points. The most important of these is that every test result has a degree of error, so results must always be treated with caution, and as only one indicator. Moreover, the reliability of any measure goes down as the number of students taking the test goes down – a basic rule of statistics. The average elementary school in Ontario has about 350 students, which means that about 40 students will write each EQAO test each year. With a group that small, scores are likely to fluctuate substantially from one year to the next. Any one score could give a highly misleading picture of performance for a typical elementary school in Ontario. Making judgments about schools based on one year's EQAO results is not a sound practice and should not be done. Ranking schools based on a single score is also dangerous. Since so many schools have very similar scores, a very small difference in results can lead to very large differences in rankings.

For the same reason, schools cannot be expected to show steady progress from year to year on these indicators. Even in an improving school or system there will inevitably be annual fluctuations, just as the daily temperature can go up or down – sometimes quite a bit! – even as the average increases as we move from winter to summer. This is another regularity of statistics; measurements rarely move smoothly in one direction.

Currently, the public analysis of Ontario results rests on only one indicator – the percent of students achieving Level 3 on the EQAO provincial tests. Focusing attention on only one indicator, which is essentially a 'bar', is potentially misleading. While the proportion of students achieving Level 3 is important, what one really wants to know is whether average performance is improving and whether the variation among students is large or small (in other words, how many students are far behind). In recent years, the proportion of Ontario students at Level 1, a very low level, has fallen by more than 50 percent. This is as important as the increase at Level 3, yet goes virtually unreported.

A true picture of the performance of any school or system requires multiple indicators over multiple years. It is reasonable to have public reporting of the achievement of public institutions, but equally important to insist that this reporting be based on enough data to give a reasonably valid picture of the situation, and that rankings should be avoided.

The Perils of Reporting the Percentage of Students At or Above Level 3

LYNNE HOLLINGSHEAD and RUTH A. CHILDS
Ontario Institute for Studies in Education, University of Toronto

In their 2010 EQAO results report, the staff of a small elementary school in Ontario's Greater Essex County District School wrote, "Our junior scores reflect our dedication to student success and the hard work we put forth. Unfortunately our primary scores do not reflect our efforts" (Centennial Central Public School, 2010, p. 1). Indeed, from the May 2009 Junior Division (Grade 6) Assessment of Reading, Writing and Mathematics to the May 2010 assessment, the percentage of students achieving at or above Level 3 (the provincial standard) decreased from 81% to 80% in Reading, but increased from 59% to 84% in Writing and from 66% to 80% in Mathematics. The Primary Division (Grade 3) Assessment, however, showed decreases in all areas between May 2009 and May 2010: 39% to 37% in Reading, 43% to 37% in Writing, and 71% to 58% in Mathematics.

This is a school's nightmare: declining results must mean that the quality of the education the school is providing is also declining, right? Not necessarily. In fact, the

Pestalozzi, by Hans Sigmund Bendal (1814-1853)

… continues on page 14

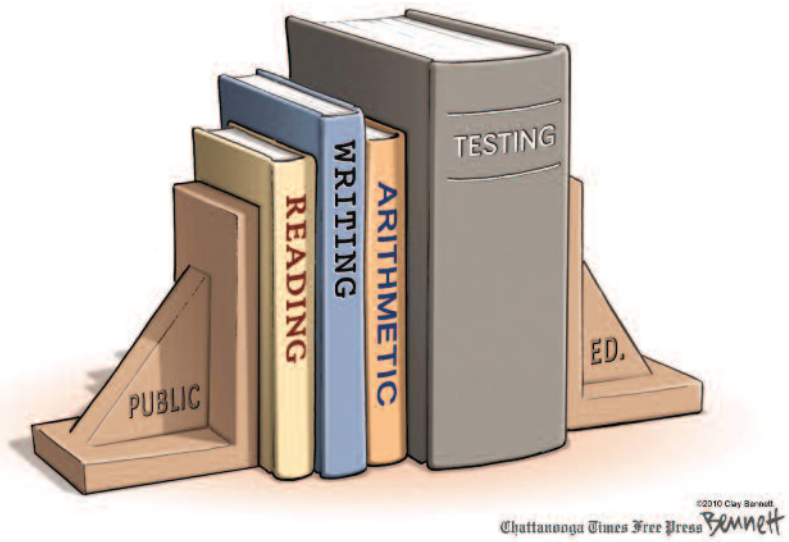
… continued from page 13

way Ontario – and many other states and provinces – has chosen to report results is not stable enough for comparing a school's results between years or for comparing results between schools or between groups of students within schools.

The Cut Score Makes a Difference

The currently common practice of reporting large-scale assessment results as the percentage of students above a pre-specified cut score (for Ontario assessments, the cut score is the border between Level 2 and Level 3) was first recommended in 1994 by an advisory committee to the state of California's testing program. As the committee noted, this type of report was attractive because it was easy to explain: the public had to understand only one level of performance – the cut score – and one number – the percentage of students above that score (Cronbach, Bradburn, & Horwitz, 1995).

Nonetheless, results reported in this way can be very misleading. Andrew Ho (2008, 2009; Ho & Haertel, 2006), building on earlier work by Paul Holland (2002), showed that small changes in the location of the cut score differentially affect schools, depending on how many students in the school have scores near the cut score. In other words, the percentage of students above a cut score is very sensitive to the location of the cut score. This problem is compounded when results are compared between schools or even, within a school, across years or between groups of students (for example, boys and girls).



Printed with the kind permission of Clay Bennett

… continues on page 15

… continued from page 14

School Size Matters

If the percentage of students above a cut score – in Ontario, the percentage of students at or above Level 3 – is an unstable index of school performance, is it worse for some schools than for others? Intuition suggests that the results would be more unstable for smaller schools. The staff at Centennial Central Public School emphasized the effect each student in their small school had on their assessment results: "In grade six, 25 students wrote this assessment (13 female and 12 male). Each student represents 4%. . . . In grade three, 19 students wrote this assessment (12 female and 7 male). Each child represents slightly over 5% points" (Centennial Central Public School, 2010, p. 1).

In our recent research (Hollingshead & Childs, 2011), we investigated the effects of school size on the stability of the percentage of students at or above Level 3 using data from the 131,119 Ontario students who had Reading results from the English-language version of the Junior Division (Grade 6) Assessment of Reading, Writing and Mathematics in 2009. We found that small schools – those with 30 or fewer students in Grade 6 – had a 33% probability of showing a difference as large as 13% in the percentage of students at or above Level 3 just by chance, when there was no real difference. As one would expect, smaller schools had less stable results and larger schools (or several schools combined) had more stable results. Comparisons within schools between groups of students – such as, girls and boys – were particularly worrying, because such comparisons were based on even smaller numbers of students in each group.

What Should We Do?

The evidence is mounting that reporting EQAO assessment results as the percentage of students achieving at or above Level 3 is misleading. It is also, we believe, unfair to small schools, such as Centennial Central, and to small school boards, which are regularly called upon to explain wildly fluctuating results.

Ho (2008, 2009; Ho & Haertel, 2006) has suggested some alternatives, including graphics that compare the percentages of students above every score, not just the cut score, and nonparametric statistics that summarize the differences between score distributions. For Ontario, we recommend *always* reporting the percentages of students at each of the four levels, and those below Level 1. This information is sometimes provided in school reports for the current year, but is not included when comparing across years or when comparing groups within years. Certainly, one could report even more detailed distributions of results (for example, students can be divided into five sublevels within each of the four main levels) for every comparison, but that may not be necessary. We recognize that even reporting the four levels (plus below Level 1) will require EQAO and others to develop new graphics and tables to summarize comparisons over time and between groups of students. While the implementation of such reports may be challenging, more stable information is

REFERENCES

Centennial Central Public School (Greater Essex County District School Board). (2010). *EQAO community report for 2009-2010.* Comber, ON: Hannigan, J. Retrieved from **http://www.gecdsb.on.ca/schools/elem/ccps/NewsletterPDFs/EQAO%20Report.pdf**

Cronbach, J. L., Bradburn, N. M., & Horwitz, D. G. (1995). Sampling and statistical procedures used in the California Learning Assessment System. In L. J. Cronbach (Ed.), *A valedictory: Reflections on 60 years in educational testing (pp. 34-36).* Washington, DC: National Academy Press. Retrieved from **http://www.nap.edu/catalog.php?record_id=9244**

Ho, A. D. (2008). The problem with "proficiency": Limitations of statistics and policy under No Child Left Behind. *Educational Researcher, 37*, 351-360.

Ho, A. D. (2009). A nonparametric framework for comparing trends and gaps across tests. *Journal of Educational and Behavioral Statistics, 34*, 201-228.

Ho, A. D., & Haertel, E. H. (2006). *Metric-free measures of test score trends and gaps with policy-relevant examples* (CSE Tech. Rep. No. 665). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards and Student Testing (CRESST). Retrieved from **http://www.cse.ucla.edu/products/reports/r665.pdf**

Holland, P. W. (2002). Two measures of change in the gaps between the CDFs of test-score distributions. *Journal of Educational and Behavioral Statistics, 27*, 3-17.

Hollingshead, L., & Childs, R. A. (2011). Percentage of students above a cut score: The effect of group size. *Educational Measurement: Issues and Practice, 30*(1), 36-43

… continues on page 16

Queen's Education Letter

Queen's Education Letter is a thematic bi-annual publication of the Education Alumni Steering Committee and Queen's Faculty of Education.

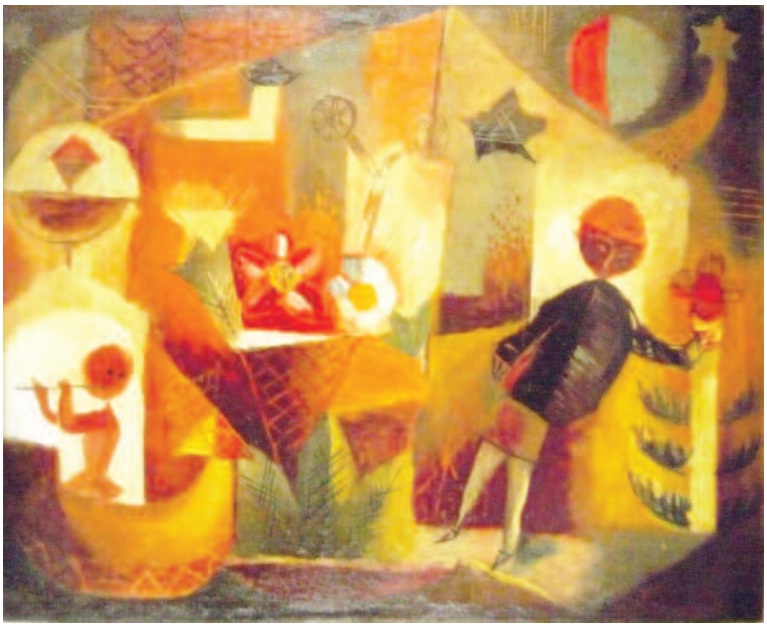
Subscription
Queen's Education Letter is currently offered free-of-charge in PDF format at www.educ.queensu.ca/alumni and will be delivered in bulk, free-of-charge, to select locations. The *Queen's Education Letter* is supported by annually-pledged donations from alumni and friends as well as matching funds from the Office of the Dean of Education.
ISBN # 978155339157-9

For questions, inquiries and comments
Carole Morrison
QUEEN'S FACULTY OF EDUCATION
Duncan McArthur Hall
Kingston, ON K7M 5R7
edletter@queensu.ca
(613) 533-6000 x75408

Queen's University
FACULTY OF EDUCATION
Duncan McArthur Hall
Queen's University
Kingston, Ontario Canada
K7M 5R7

… continued from page 15

needed to support Ontario education and minimize the misinterpretation of the results that are obtained. Based on their report, the teachers and principal at Centennial Central state they are "working diligently to improve the primary results" (Centennial Central Public School, 2010, p. 1). Until we are able to improve the quality of the reports our schools receive and the stability of the information underlying these reports, we cannot assure educators that their efforts are well-founded.



Schoolbook of Art, by Koga Harue, 1926