A New Vision for

Large-Scale Testing in Ontario









December 2011











We hope to work with the Ministry of Education and EQAO to revise the current provincial testing model to better serve the needs of students, teachers and schools through more effective accountability measures that are supported by the most current and sound educational research.

A New Vision for Large-**Scale Testing in Ontario**

INTRODUCTION

Ontario's publicly funded education system has made remarkable progress since the current Liberal Government was elected in 2003. Teachers, educational workers, parents and students appreciate the intense interest in improving our education system. Premier McGuinty has earned a reputation as one of the preeminent leaders in educational innovations in the world. Our education system has moved in a positive direction and it continues to improve to meet the needs of our students and the expectations of the public through initiatives aimed at:

- reducing primary class size;
- renewing the focus on professional development;
- recognizing the importance of literacy and numeracy;
- targeting resources and supports to the most vulnerable schools;
- focusing on supporting disengaged students at the secondary level;
- increasing graduation rates;
- enhancing supports for students with special needs;
- reducing gaps in student achievement.

While many of these programs and initiatives within the publicly funded education system have been effective, there is a need to examine the current provincial testing model. We are requesting a review of the mandate of the Education Quality and Assessment Office (EQAO).

BACKGROUND INFORMATION

The original mandate of EQAO came out of the Royal Commission on Learning (RCL) in December 1994. The main focus was to monitor the curriculum and how well students were learning it. Success was to be determined by use of a uniform assessment for all students at the end of Grade 3. As a small agency, the Office of Learning Assessment and Accountability (OLAA), was to be responsible for the administration, scoring and reporting of these assessments.

The RCL enunciated its philosophy and vision for the OLAA in Chapters 11 and 19:

"We believe that some systemwide testing should be built in, as a check on student learning at a few critical transition points, and as a vehicle for assuring people that, at those points, all students are being assessed according to the same yardstick."

"In addition, the tests are to enhance reporting to parents and for examining the content and delivery of curriculum. Test results are, most emphatically, not to be used to place or sort students for any reason. They will serve as the central check on how effectively the curriculum is serving the learning needs of the students..."

CURRENT STATE

The EQAO's mandate has strayed significantly from the original vision for the OLAA as set out by the RCL, which envisioned:

- sample-based assessment, not universal, extensive and expensive testing;
- tests that would act as a check on the effectiveness of the curriculum, not as a means of sorting students, or ranking schools - for example, "shopping" for schools;
- a review of the work and mandate of the OLAA after five years in the expectation that its work might prove redundant rather than requiring the addition of census tests in Grades 6 and 9.

As well, the many initiatives and changes in practice and policy that have occurred since the implementation of the EQAO necessitate a rethinking of the current model of large-scale assessment in Ontario. When the OLAA was conceived, there was virtually no standardization in Ontario education. In the ensuing years, the Ontario Ministry of Education introduced a standardized report card and standardized performance levels in the form of achievement charts in all curriculum documents. It developed a policy document, entitled Growing Success, which aims at standardizing assessment, evaluation and reporting across Ontario. During the same period, significant research on the importance of formative assessment conducted by the teacher in the classroom was brought to the attention of the education community.¹ The Ministry of Education and the EQAO have both touted the importance of this formative assessment research. Extensive professional development focusing on standardizing teacher performance in conducting formative assessment in literacy and numeracy has been implemented at both the provincial and local levels through the Literacy and Numeracy Secretariat and Student Success initiatives. The assessment context in Ontario, now significantly different from that of the OLAA era, makes many features of the EQAO testing redundant and regressive.

Both the PIRLS and PISA assessments, which are based on random sampling, have earned world-wide legitimacy as indicators of the effectiveness of a school system. This format, both statistically reliable and valid, would provide sufficient data for making decisions about curriculum and generating public confidence that our system is effective and accountable.

ISSUES

1) The purpose for data collection should drive decisions about provincial testing

EQAO's mandate is to ensure greater accountability and to enhance the quality of education in Ontario. This public accountability mandate can be fulfilled through random testing. Such a testing program provides sufficient data both to ensure that Ontario students are meeting literacy and numeracy expectations and to develop and implement changes to curriculum.

2) Duplication of data

Ontario's students already participate in national and international assessments that analyze their performance in Reading, Mathematics, and Science. These assessments, through random sampling of students, provide reliable metrics on student achievement both in Ontario and globally.

The 2010 Pan-Canadian Assessment Program (PCAP) results provide additional evidence to support both decision-making about curriculum and public confidence in the system.

The most recent Progress in International Reading Literacy Study (PIRLS) and Programme for International Student Assessment (PISA) ² results indicate that Ontario has made huge strides in raising student achievement levels in our elementary and secondary schools. The public can remain confident that Ontario schools provide education equal to that of the highest achieving countries.

In the 2009 PISA results, Ontario was among the top-performing jurisdictions in the world. Only Shanghai-China performed better than Ontario in Reading, the major focus of the 2009 PISA. Of the 75 participating jurisdictions, six were statistically the same as Ontario, including British Columbia and Alberta. Ontario students performed at or above the Canadian average in Reading, Mathematics and Science. In the 2009 results, while students in many Canadian provinces recorded significant declines in achievement, Ontario students maintained high performance

² PIRLS is an international testing program initiated by the International Association for the evaluation of Educational Achievement of Grade 4 students. PISA is an international program initiated by the Organization for Economic Co-operation and Development (OECD) and undertaken every three years to assess the achievement of 15-year-olds in three domains of Reading, Mathematics and Science.

Research done internationally and in Ontario indicates that formative assessment is the best form of assessment. Formative assessment provides immediate feedback to the child and the parents and ensures that expectations are learned well and for long-term sustainability.

levels. This is indicative of effective teaching and learning programs in place across Ontario. 3

Higher standards and closing the achievement gap are common aspirations among Ontario teachers. The PISA results indicate that standards are already high in Ontario and that the achievement gap is closing. Another indicator of our high level of success on the PISA is significantly fewer Ontario students with test scores in the lowest levels. Ontario is a world leader in closing the gap between the lowest and highest achievement levels. Although many recent policy initiatives in Ontario emulate those of Great Britain, Australia and the United States, those countries could match neither Ontario's achievement in the international rankings nor its significant reduction in the gap between highest and lowest achievers. The Organization for Economic Cooperation and Development (OECD) attributes this reduced gap to Canadian social policies addressing economic and social inequities.4

Additional data from PISA indicate that Ontario is one of the highest achieving jurisdictions in the world with highly educated and effective teachers.⁵

Results from the 2010 PCAP show that Ontario is the only Canadian jurisdiction whose students are above the national average in all three subject areas – Reading, Mathematics and Science. Further, results indicate that:

- Ontario was the only jurisdiction above the Canadian average in Reading;
- Ontario was one of the two provinces whose students were above the Canadian average in Math. Of these 92% achieved at Level 2 (considered the expected standard) or higher.
- Ontario was one of two provinces whose students were above the Canadian average in Science. 6

⁴ OECD, PISA

⁵ OECD, PISA

⁶ EQAO 2011

Significant evidence exists that classroom-based formative assessment is a more reliable indicator of individual achievement and most *closely aligned with the best strategies* for individual student improvement. It is also the best predictor of student achievement.

3) Research has proven that formative assessment (assessment for/as learning) more effectively promotes student learning than summative assessment (assessment of learning).

Assessment for learning is continuous. It is used to monitor student performance, provide timely feedback and generally enhance the teaching-learning experience. It occurs during instruction to support the next stage of learning.

Through Ministry of Education initiatives and excellent professional development (PD) opportunities, teachers have expanded their expertise in quality assessment and linked assessment with instruction. Ontario teachers have learned more about assessment for/as learning through the implementation of Growing Success. Teachers need more opportunities to learn about effective assessment to inform their practice. Ministry investment in PD, to help teachers develop their formative assessment skills, would be money well spent. EQAO's own website and publications concur that this type of assessment is the most important component of student learning.⁷

The RCL (1994) supported this belief in the importance of teacher assessment versus large-scale assessment. Some of its comments in this regard are shared below:

"Learning does not proceed in neat steps...; therefore, tests cannot be applied to students as simply as quality control can be applied to objects coming off a conveyor belt. Tests will not fix students' problems or improve teaching..... At best, they can tell parents something (but never everything) about what their children know...."

"Evaluating students regularly enables teachers to monitor learning and make changes when learning is not occurring in sufficient depth. Regular evaluations... assure students that they understand what is being taught and can move onto the next task, thus advancing student learning. We call this formative evaluation, because it helps form the learning and teaching needed to achieve success."8

The publicly funded education system must be accountable to the public. Stakeholders in the education system must work together to provide information that is both relevant and contextualized to ensure public confidence in the system.

"The assessments that drive student learning and academic self-concept are those used by teachers in classrooms. Without quality classroom assessment, instruction cannot work and schools cannot be effective."9

The RCL emphasized formative assessment as a key to promoting learning. It did not envision that the implementation of large-scale testing would eclipse that role by shifting time and attention toward summative assessment in the form of EOAO tests.

4) Distorted value is placed on snapshot-in-time testing results provided by EQAO

We are well aware that the province requires data to make educational policy decisions and to guide educational initiatives. However, data obtained from snapshot-in-time assessments should not form the basis of policy decision-making as it does currently. Measurement specialists oppose high-stakes testing because using a single indicator of competence to make important decisions about individuals or schools violates the professional standards of the measurement community.¹⁰

"...it is important to emphasize that large-scale testing has limitations; otherwise, people reach what we are convinced is the mistaken conclusion that these few tests are the most important in the student's school career, or that many such tests would be ideal. In our opinion, large-scale testing is unlikely to be a more fair and accurate representation of student learning than the best judgment of the well-trained teacher-assessor. Moreover, such testing is easily misused." 11

"The public has been brainwashed into believing that testing is a means to ensure educational standards and an objective basis on which to distribute educational resources – despite the lack of compelling evidence that these alleged benefits outweigh the considerable demonstrated disadvantages." 12

⁹ Stiggins, R. (2002), p. 760

¹⁰ Nichols, Sharon L., Glass, Gene V, Berliner, David C., (2006)

¹¹ Royal Commission on Learning (1994), p. 138

¹² Popham, W. (1999), p. 13-17

Data about student achievement obtained from alternative measures/means should be as relevant as the data obtained from large-scale assessment. Report cards are an important source of data for students and parents since they now include the number of late periods, absences, final marks, as well as the learning skills and work habits for each subject.

Contextual factors such as access to resources, social and economic status, cultural and family background as well as issues of health and safety are significant drivers of individual student success. For that reason, the Ministry of Education has successfully implemented initiatives to close the gap between students with much social capital and those with little.

"We have known for more than three decades that family background factors alone explain in excess of 50% of the variation in student achievement. This widespread and unnecessary practice [school rankings] demeans the work of schools serving less advantaged students, masks the significant contributions made by many of these schools to the growth of their students and lionizes the work of schools serving exceptionally advantaged students, even if they are not adding much of value." ¹³

The RCL (1994) also addressed concerns about allowing data to be used in a manner that might have negative effects on students and the system.

"Any single test used for large-scale assessment and reporting assumes a distorted importance and can – and often does – have long-term, frequently negative consequences for students and for the learning system, because of the inappropriate ways the information is used." 14

Further stigmatizing school communities, the School Information Finder enables parents to "shop" for schools based on rankings done according to EQAO results. Such rankings or those produced by organizations such as the Fraser Institute often form the basis upon which parents select a school for their child.

5) EQAO testing impinges on classroom instruction

Many people falsely assume that the EQAO assessments monitor the quality of instruction. But there is an enormous difference between instruction for improving student learning and instruction for improving student test scores. Improving student learning requires thoughtful instructional strategies, significant

School-based assessments, strategies and tools can provide teachers, schools and boards with pertinent formal information about how students are learning in reading, writing and mathematics in a less intrusive way.

time and a variety of resources. The annual EQAO testing and the test preparation take important time away from regular classroom instruction. Teachers are forced to compromise good pedagogy as they respond to external pressures to spend significant instructional time preparing for the test.

"Every hour spent on such exam preparation is an hour not spent helping students to become critical creative curious learners." 15

"...pressure of doing well on a test seriously compromises instructional practices." 16

The EQAO document, Summary of Results and Strategies for *Teachers, 2007-2008* also notes a compromise in instructional practices in the recent movement towards more formulaic responses by students. They state that "over the last few years, scorers have noted increasing student use of formulas to improve the quality of answers to open-response reading questions. Most of these formulas help the students to construct answers that use words from the question, respond to the question, provide proof from the text and make a personal connection in conclusion. Scorers have noted that many students have learned these formulas but apply them mechanically (i.e., with little relationship to the text or the question asked)."17

Many good practices have been developed by the LNS, but teachers and the system need time to assimilate and apply them (e.g., teacher moderation). The Ministry should provide school boards funding for resources and time to implement interventions to further narrow the achievement gap. The resources needed should be found through a reduction in the time and money currently spent on EQAO testing.

6) EQAO testing is not beneficial for students

Results from EQAO assessments arrive when students have moved on to the next grade. While the format of the Individual Student Report (ISR) has improved, there is little context for the student. For Mathematics, especially, the learning benefit for

EQAO assessments measure a very narrow span of a child's potential. They cannot reveal the many other areas in which students can excel such as the arts, physical education and technology.

the child is seeing where he/she has made errors. This cannot happen with the ISR. The feedback included on the new ISR under "Next Steps" is very general and has less learning significance when received several months later. Additionally, as shown below, the format of the ISR does not provide an analysis of the child's individual strengths and weaknesses. ¹⁸

YOUR CHILD'S RESULTS ON THE JUNIOR DIVISION ASSESSMENT (GRADE 6, 2011)

	NE1 Not enough evidence to be assigned a Level 1	LEVEL 1 Much below provincial standard	LEVEL 2 Approaches provincial standard	LEVEL 3 Provincial standard	LEVEL 4 Exceeds provincial standard
READING: answered questions: 36 of 36				-	
WRITING: answered questions: 14 of 14			-		
MATHEMATICS: answered questions: 36 of 36				•	

FOR PARTICIPATING STUDENTS, THE SYMBOL REPRESENTS THE STUDENT'S ACHIEVEMENT LEVEL ON THIS ASSESSMENT.

[&]quot;Answered questions" refers to the number of questions the student answered in relation to the total number of questions on the junior division assessment. Unanswered questions are treated as incorrect in the calculation of your child's results.

Individual student assessment should be the responsibility of the classroom teacher, supported by school board and Ministry in-service and resources but EQAO testing can and often does interfere with the capacity of teachers to focus on high-quality classroom assessment.

"We ought to be providing environments that enable each youngster in our schools to find a place in the educational sun. But when we narrow the program so that there is only a limited array of areas in which assessment occurs and performance is honored, youngsters whose aptitudes and interests lie elsewhere are going to be marginalized in our schools." 19

Motivation to learn is the most powerful predictor of student achievement. EQAO tests originally set an achievement standard at level two, which was later arbitrarily changed by the Conservative Government to level three. International testing agencies have always set the standard at level two. Students who may have made encouraging gains and have reached level two, may still feel they do not measure up and their self concept, which is related to their motivation to learn, will be damaged. ²⁰

Every teacher knows the enormous effort it takes_to struggle from level one to level three, especially for students with substantial learning difficulties. Low test scores often lead to a self-fulfilling prophecy on continuing lower achievement. ²¹

"... the apparent rise in eating disorders, burn-out, male disaffection and behaviour problems, even in the more academic schools, are signs that all is not well. ... there is an unhealthy and neurotic perfectionism developing among children which involves a strong desire to avoid mistakes."²²

The recent People for Education report states that the number of students who read for enjoyment has dropped significantly over the last ten years. The data comes from the student questionnaires that accompany the Primary and Junior Assessments of Reading, Writing and Mathematics.

The report cites several possible reasons for the decline in numbers. Of particular concern is the intense focus on literacy which causes students to see reading as work and not as enjoyment. ²³

Finally, the current misuse of large-scale assessment results to compare and rank students, schools and districts does not improve student learning.

¹⁹ Eisner, E. (2001). p. 367

²⁰ Paris, S., Lawton, T., Turner, J. & Roth, J. (1991), p. 12-20

²¹ Koretz, D. et al. (1991)

²² U.K. Guardian, November 22, 2001

²³ People for Education, December 12, 2011

Ontario teachers are committed to working with the Government in order to move forward with positive reforms to the province's testing model. We offer to work in partnership to promote these reforms to parents and the broader public.

CONCLUSION

Our belief is that the purpose of large-scale testing should be to ascertain how well students are learning Ontario's curriculum. Ontario's students rank highly on international tests, making largescale assessments, for all Grades 3, 6 and 9 students every year, unnecessary in our present educational context.

Efficient alternative ways of collecting data to make educational policy decisions that are less intrusive, time-consuming and costly exist. We believe it is time for Ontario to move away from its reliance on EQAO tests and to develop alternative means of data collection.

We also believe that reducing the amount of EQAO testing will have the following effects on the education system as a whole:

- increasing student enjoyment of reading and writing;
- freeing instructional time for other curriculum;
- refocusing student success in a more rounded manner;
- maintaining the connection with curriculum and the curriculum review process;
- decreasing stress across the system;
- reducing costs for test administration and providing more funds for assessment resources;
- increasing assessment flexibility for making educational decisions;
- making feedback more timely; and
- reducing the media and commercial exploitation of results.

If Ontario is to achieve its goal of helping all students be successful learners and reach their full potential, the province should focus more on what the research and the professionals in the classroom are saying about how best to assess and support student learning. Like the Government, Ontario teachers and educational workers want what is best for students and our publicly funded schools.

RECOMMENDATIONS

- 1. That the Ministry of Education encourage EQAO to work with OTF and the Affiliates to review the current model of large-scale testing in light of the issues raised in this paper.
- 2. That the Ministry of Education, in partnership with EQAO, OTF and the Affiliates, collaborate to explore alternatives to provincial large-scale testing policies and methods.
- 3. That the funds saved through a change from census to random sample testing and a reduction in the size of the EQAO be redirected into the education system.

BIBLIOGRAPHY

American Educational Research Association (1999).

Standards for educational and psychological testing. Washington, DC: Author.

Amrein, A. L. & Berliner, D.C. (2002a).

The impact of high-stakes tests on student academic performance: An analysis of NAEP results in states with high-stakes tests and ACT, SAT and AP Test results in states with high school graduation exams. Tempe, AZ: Education Policy Studies Laboratory, Arizona State University. Retrieved June 7, 2008 from http://www.asu.edu/educ/ epsl/EPRU/documents/EPSL-0211-126-EPRU.pdf

Amrein, A. L. & Berliner, D.C. (2002b).

High-Stakes testing, uncertainty and student learning Education Policy Analysis Archives, 10(18). Retrieved June 7, 2008, from http://epaa.asu.edu/epaa/v10n18/

Black, P., & Wiliam, D., (1998a).

Assessment and classroom learning. Assessment in Education: Principles, Policy and Practice, 5(1), 7-73. http:// www.youtube.com/watch?v=xeS6U-eDkKA

Black, P., & Wiliam, D. (1998b, October).

Inside the black box: Raising standards through classroom assessment. Phi Delta Kappan, 80(2), 139-149.

Education Quality and Accountability Office (November, 2007).

News Release. Ontario Grade 4 Students' Reading Abilities Among the Best in the World: Elementary school students shine among Canadian provinces and internationally. Retrieved June 1, 2008 from EQAO Web site: http://www. egao.com

Education Quality and Accountability Office (November, 2007).

Programme for International Student Assessment (PISA), 2006 Highlights of Ontario Results. Retrieved December 1, 2007 from the EQAO Web site: http://www.eqao.com

Education Quality and Accountability Office (2009).

Summary of Results and Strategies for Teachers 2007-2008. Queen's Printer for Ontario.

Eisner, E. (2001).

"What Does It Mean to Say a School Is Doing Well?" Phi Delta Kappan (January): 367-72.

Fullan, M., Hill, P. & Crevola, C. (2006) Breakthrough. Corwin Press. Hess, F. (Ed.), (2005). Urban school reform: Lessons from San Diego. Cambridge, MA: Harvard Education Press.

Guardian, U.K., November 22, 2001.

"Success, the new drug".

International Association for the Evaluation of Educational Achievement (IEA) (2007).

Progress in International Reading Literacy Study (PIRLS) co-ordinated in Ontario by the Education Quality and Accountability Office (EQAO).

Jones, M. G., Jones, B. D. & Hargrove, T. (2003).

The unintended consequences of high-stakes testing. Lanham, MD: Rowman & Littlefield.

Kohn, Alfie, (2000).

The Case Against Standardized Testing: Raising Scores, Ruining the Schools, (Heinman Portsmouth, NH.

Kohn, Alfie, (2001).

Fighting the Tests A Practical Guide to Rescuing Our Schools. Phi Delta Kappan, January. Retrieved June 7, 2008, from http://www.alfiekohn.org/ teaching/ftt.htm

Koretz, D., Linn, R. L., Dunbar, S. B., & Shepard, L. A. (1991). The Effects of High-Stakes Testing: Preliminary Evidence About Generalization Across Tests. In R.L. Linn (Chair), The Effects of High Stakes Testing Symposium presented at the annual meeting of the American Educational Research Association and the National Council on Measurement in Education, Chicago.

Leithwood, K. (2001).

"Five Reasons Most Accountability Policies Don't Work (and What You Can Do About It)." Orbit, 32(1), p. 4.

Marlaire, C. & Maynard, D., (1995).

"Standardized Testing as an International Phenomenon," Sociology of Education, 63, 83-101.

Michigan Association of School Psychologists, (2004).

Position statement on the use of the Michigan Educational Assessment Program (MEAP) for high stakes decisions for students in Michigan. Retrieved January 2, 2005, from the MASP Web site: http://www.masponline.org/HighStakes_Testing_in_Michigan.htm.

Moll, M. (2004).

Passing the Test. Ottawa. Canadian Centre for Policy Alternatives and the authors.

Nichols, S. L., Glass, G. V. & Berliner, D. C. (2006).

High-stakes testing and student achievement: Does accountability pressure increase student learning? Education Policy Analysis Archives, 14(1). Retrieved June, 2008 from http://epaa.asu.edu/epaa/v14n1/ The International Association for the Evaluation of Educational Achievement (IEA), 2006 Progress in International Reading literacy Study (PIRLS).

Noddings, N. (2001).

Care and coercion in school reform. Journal of Educational Change, 2, 35-43.

Noddings, N. (2002).

High-stakes testing and the distortion of care. In J. L. Paul, C. D. Lavely, A. Cranston-Gringas & E. L. Taylor (Eds.), Rethinking professional issues in special education (pp. 69-82). Westport, CT:Ablex Publishing Corporation.

Organisation for Economic Co-operation and Development (2006).

Program for International Student Assessment (PISA) – Co-ordinated in Canada by Human Resources Development Canada (HRDC), Council of Ministers of Education Canada (CMEC), Statistics Canada, and provincial ministries and departments of education.

Paris, S., Lawton, T. Turner, J. & Roth, J. (1991).

"A Developmental Perspective on Standardized Achievement Testing," Education Research 20, 12-20.

Pedulla, J. J., Abrams, L. M.E., Madaus, G. F., Russell, M. K., Ramos, M.A. & Miao, J. (2003).

Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers. Boston, MA: Boston College, National Board on Educational Testing and Public Policy. Retrieved June 7, 2008 from http://www.bc.edu/research/nbetpp/statements/nbr2.pdf

People for Education (2011)

Reading for Joy. Report from People for Education, December 2011.

Popham, W. (1999).

"Where Large Scale Educational Assessment is Heading and Why It Shouldn't," Educational Measurement: Issues and Practices, 13-17. The Effects of Important Tests on Students: Implications for a National Examination or System of Examinations. Paper presented at the American Education Reform Association Conference, Washington, DC.

Popham, W. J. (2008).

Transformative Assessment. Alexandria, VA: Association for Supervision and Curriculum Development.

Royal Commission on Learning (1994).

For the Love of Learning. Report of the Royal Commission on Learning. Queen's Printer for Ontario.

Ryan, J. E. (2004).

The perverse incentives of the No Child Left Behind Act. New York University Law Review, 79, 932-989.

Spatig, L. (1996).

Developmentalism Meets Standardized Testing. Paper presented at the annual meeting of the American Educational Studies Association, Montreal.

Stiggins, R. (2002)

Assessment Crisis: The Absence of Assessment for Learning, Phi Delta Kappan, June 2002, pp. 758-765.

ACKNOWLEDGEMENTS

Thanks are due to many who have contributed to this position paper. The Ontario Teachers' Federation and its Affiliates wish to extend a special thank you to Nanci Wakeman for her expertise in editing and her knowledge of assessment practices.