

Ricci, C., & Taylor. S. (2004, Fall). Challenging the validity of standardized testing. *Our Schools/Our Selves*, 14(1) #77, 63-73.

V. 14 N.1 (#77) Fall 2004 \$12.00

OUR SCHOOLS

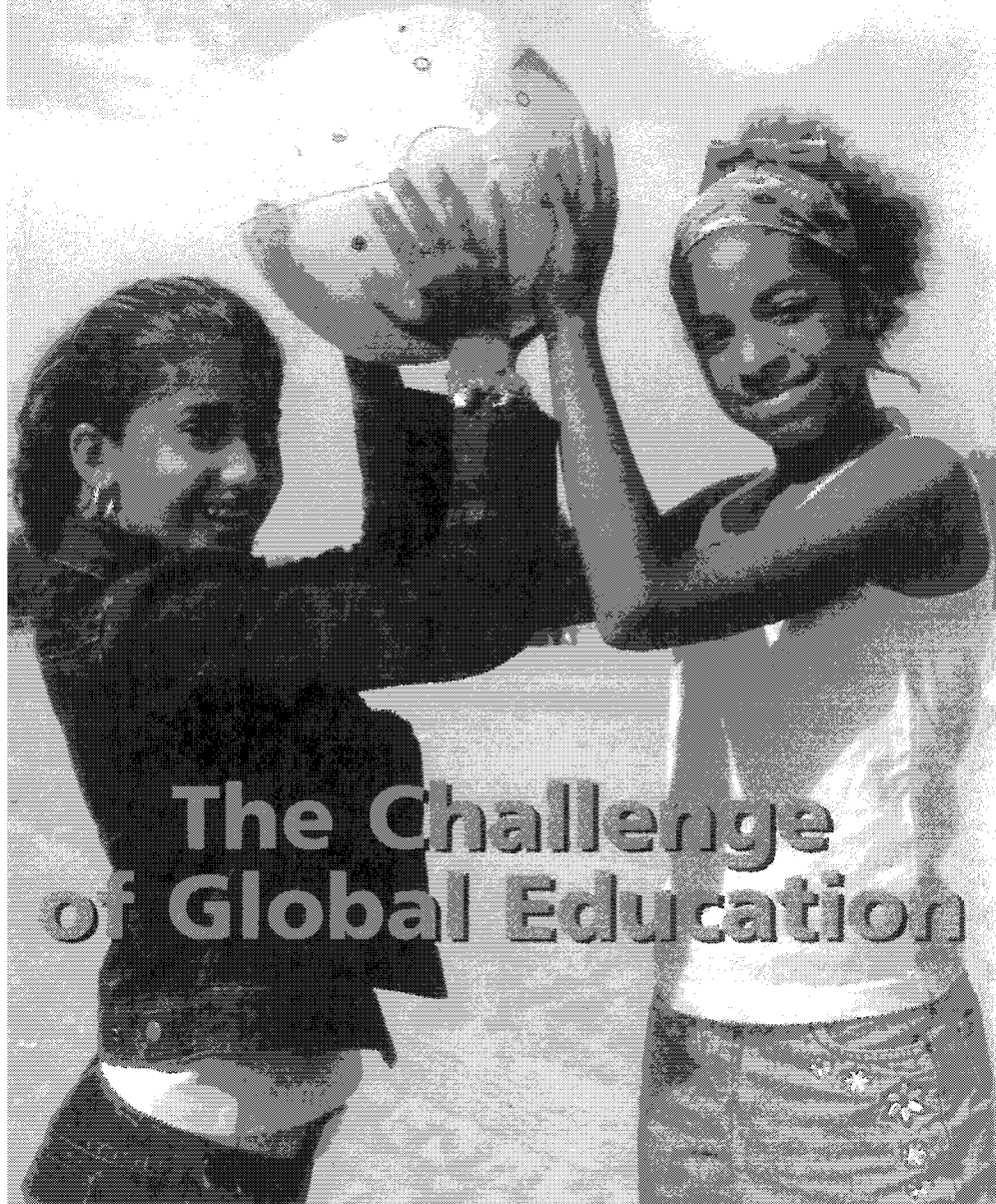
The Canadian Centre for Policy Alternatives

OUR SELVES

How valid are
standardized tests?

The many faces
of privatization in
Canada's public schools

A native story teller in
an inner city classroom



The Challenge of Global Education



Our Schools/Our Selves is published four times a year by the Canadian Centre for Policy Alternatives, 410-75 Albert St., Ottawa, ON K1P 5E7. This is Volume 14, Number 1, Issue #77 of the journal (Fall 2004). Our Schools/Our Selves is a member of the Canadian Magazine Publishers Association. It is indexed in the Canadian Magazine Index and the Alternative Press Index.

Editorial Board

Keren Brathwaite, Denise Doherty-Delorme, Bernie Froese-Germain, Larry Kuehn, Doug Little, George Martell, Patricia McAdie, Marita Moll, Satu Repo, Heather-jane Robertson, Erika Shaker

Advisory Board

Lynn Carlile, Kari Dehli, Sarah Dopp, Ed Finn, David Flower, Karin Jordan, Clare Keating, Kirsten Kozolanka, David Livingstone, Mike Luff, Diane Meaghan, Janice Newson, Aleda O'Connor, Claire Polster, Anne Rodrigue, David Robinson, Mark Rosenfeld, Harry Smaller, Jennifer Sumner, Jim Turk, Sofia Vuorinen, Robert Whiteley

Editors

Satu Repo and Erika Shaker

Editorial Office

107 Earl Grey Road, Toronto, ON M4J 3L6, tel or fax (416) 463-6978

Subscriptions and Advertising

Canadian Centre for Policy Alternatives, 410-75 Albert St., Ottawa, ON K1P 5E7, tel: (613) 563-1341 fax: (613) 233-1458

ISSN 0840-7339

Production

Typesetting and design: Nancy Reid. Printed in Canada by Imprimerie Gagne, 80 Ave. St. Martin, Louiseville, PQ J5V 1B4. Publications Mail Registration No. 8010.

Cover Design

Nancy Reid

Front Cover and Page One Photo

Nancy Reid

Special thanks to Alyssa, Coral, Emma, Ethan, Isabella and Sofia for lending their smiles to the cover and page one photos.

Contents

Volume 14, Number 1 (#77)

Fall 2004

- 9 Education Roundup**
Larry Kuehn
- 19 Editorial Notes**
- 19 Why are school libraries in decline?**
Satu Repo
- 23 Litigating literacy**
Heather-jane Robertson
- 29 Viewpoints**
- 29 Standardized testing in Ontario Schools**
Bob Davis
- 33 Challenging corporately co-opted schools**
Jessica Lax and Jocelyn Land-Murphy
- 37 From Other Media**
- 37 No Politician Left Behind**
Deborah Meier
- 40 The class issue Labour has dodged**
Fiona Miller
- 43 Global Education and good citizenship**
Melanie Young

-
- 53 Why we must remember the Rwandan genocide**
Patricia Froese-Germain
- 63 Challenging the validity of standardized testing**
Carlo Ricci and Steven Taylor
- 74 Hurry, hurry**
A poem by Clydia Forehand
- 77 Getting a read on education privatization**
Bernie Froese-Germain
- 87 The trouble with accountability contracts**
Robert Whiteley
- 97 The Business of School Inc. in B.C.**
Joni Miller
- 119 The life of Rodney Bobiwash, a native educator and activist**
Jason Price
- 131 Talking to children in classrooms: myth-telling as truth-telling becomes social justice learning**
M.C.L. Provost-Turchetti

-
- 141 The University Front**
141 Life or Debt
Erika Shaker and Kerry-Anne Finn
- 145 Why the future looks bleak for Nova Scotia students**
Darcy Knoll
- 149 Media Literacy**
Michael Moore's Candid Camera
A review by Frank Rich
- 155 Book Review**
Weaving Connections: Educating for Peace, Social and Environmental Justice, edited by Tara Goldstein and David Selby
A review by Myra Novogrodski



Teaching. It's about the whole child.

Ontario's 65,000 public elementary school teachers and education workers challenge our children's imagination, open their minds, and celebrate our diversity.

**We're working with parents
to help every child succeed.**

ETFO FEEO
Elementary Teachers' Federation of Ontario
Fédération des enseignantes et des enseignants
de l'Élémentaire de l'Ontario

www.etfo.ca

Challenging the validity of standardized testing

Examinations are formidable even to the best prepared, for the greatest fool may ask more than the wisest man can answer.

— Charles Caleb Colton, author (1780-1832)

Abstract

In 1996 the Ontario Ministry of Education created the Education Quality and Accountability Office (EQAO). As of 1999-2000 all high school students in Ontario are required to pass the Education Quality and Accountability Office's Ontario Secondary School Literacy Test (OSSLT) (or now a grade 12 equivalency course) before they are eligible to graduate from high school. The Minister of Education issued a policy entitled, "Ontario Secondary Schools, Grades 9 to 12: Program and Diploma Requirements, 1999." The policy states the following:

All students who entered Grade 9 in the 1999-2000 school year or in subsequent years must successfully complete the provincial

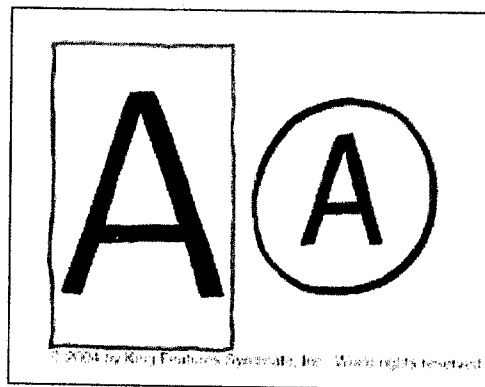
secondary school literacy test in order to earn a standard school diploma. (The provincial secondary school test, 1999)

The Grade 12 Literacy Course demands the same level of literacy as the Grade 10 Literacy Test. On April 24, 2003, the Ministry of Education issued a back-grounder entitled, "Helping students improve their literacy and math skills." It states the following:

In addition, a comprehensive, full credit literacy course, which is equivalent to the components of the Ontario Secondary School Literacy Test, will be prepared for students in Grade 12 who have failed the test. The goal is to provide them with the necessary skills to meet the requirements of the Ontario Secondary School Diploma. The course is designed to help students who have not had the full benefit of the new curriculum that includes a strong focus on reading and writing in the early years. Students must pass the literacy course to receive their Ontario Secondary School Diploma.

(Helping students improve their literacy and math skills, 2003)

To date 120,000 students have failed the test. The arms-length government organization that marks the test assures the public that the inferences they make based on this test are valid. For this project we are going to challenge those inferences. In *Introduction to Research in Education*, sixth edition (2002) Donald Ary, Lucy Cheser Jacobs, and Asghar Razavieh point out that "Validity is the most important consideration in developing and evaluating measuring instruments" (p. 242). The new *Standards for Educational and Psychological Testing* (1999), prepared by the American Educational Research Association (AERA), defines validi-

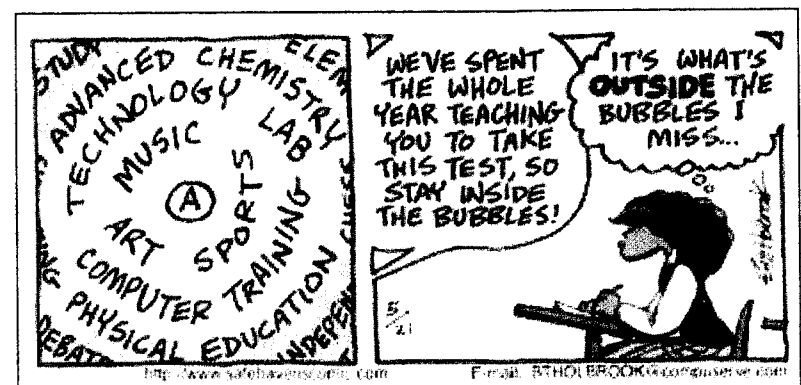


ty as "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (p. 9).

Although our research focuses on criterion-validity, in challenging the inferences made by EQAO we also consider construct and predictive validity.

To challenge criterion-validity we have asked if the test scores match other possible indicators of student literacy. What we have done is to have each grade 10 teacher within our chosen school who is willing to volunteer, review their class lists and identify those students that they believe are not literate enough to earn an Ontario Secondary School Diploma (OSSD). Then we compared this information with the EQAO OSSLT results. The EQAO OSSLT infers that all students who fail the test are not literate enough to deserve a high school diploma. If the inference is valid then another measure (in our case teachers' judgments) should agree that a particular student is not literate enough to earn an Ontario Secondary School Diploma; however, if the teacher and the OSSLT results do not agree, then the claim that the EQAO's inference is valid is flawed. This is especially significant since the OSSLT is a high-stakes test that students must pass before they can graduate.

We found that teachers disagreed with the OSSLT results identifying who should fail and who should pass, and we also discovered that teachers themselves had differing opinions about which students are not literate enough to deserve a high school diploma.



Introduction

Tests are not valid in themselves, but the inferences made about tests are either valid or not. Popham (1999) reminds us that, “Tests, themselves, do not possess validity” (p. 42). In speaking of the OSSLT we cannot, therefore, argue that the test in itself is valid or not. What we need to do is consider the inferences that are made as a result of the test. The EQAO and the Ontario Ministry of Education claim that those who fail the test are not literate enough to earn an OSSD. Popham (1999) goes on to say that “the focus of validity should be on test-based inferences and not on tests themselves” (p.44). The question then is whether the inferences made by the EQAO and the Ontario Ministry of Education about students’ literacy are valid. In challenging whether the test’s inferences are valid, we need to consider construct validity, predictive validity, and criterion validity.

Construct Validity

Literacy, like anxiety and creativity, is a construct. A construct is an abstract concept derived from a theory or from observation. The OSSLT claims to accurately measure whether a student is literate enough to earn an OSSD. They do this by using a concrete tool (the OSSLT) to measure an abstract construct. Ary, Jacobs, and Razavieh (2002) discuss construct under-representation (p. 243). Construct under-representation “refers to assessment that is too narrow and fails to include important dimensions of the construct. The test may not accurately sample some kinds of content or some types of responses or psychological processes and thus fails to adequately represent the theoretical domain of the construct” (p. 243). For example, when Dr. Ricci was teaching a grade 9 applied level course (In Ontario, courses are divided into academic, applied and essential) at the secondary school he found that students that had trouble reading canonical works would thrive when reading student selected materials. In one case, a student was reading a skateboarding magazine and Dr. Ricci needed the assistance of the student to help make sense of some of the technical terms that are exclusive to skateboarding. Similarly, most of us who are not lawyers, if asked to

read a legal document would need the assistance of a lawyer to decipher the text. This clearly suggests that a student’s literacy is contingent on the items chosen by the test developers. These are capricious decisions that build tests but do not decisively assess whether a student is literate enough to earn an OSSD.

Another aspect of construct validity is construct-irrelevant variance. This “refers to the extent to which test scores are affected by processes that are extraneous to the construct” (Ary et al., 2002, p. 243). For example, 66% of students for whom English is a second language (ESL) failed the OSSLT in the 2002-2003 school year. Does this failure mean that those students are not literate enough to earn an OSSD, or is this a product of a test that is written by and for people whose only language is English? In this case the process that has caused failure is not the process that is being assessed. Is English the only language of literacy? Does this not discriminate unfairly against people who contribute to our society despite their difficulty with English? This test is supposed to measure literacy and therefore the test takers facility with spoken and written English is a construct-irrelevant variance. In our view this amounts to racism. We need to make a distinction between language and literacy. A person who speaks another language may be literate despite their lack of English fluency.

Predictive Validity

Predictive validity is a measure of how well test scores correlate or predict a particular outcome. For the purposes of this study, this means that if a student excels on the OSSLT then we should be able to infer or predict their success in later coursework. Likewise, if a student scores poorly on the OSSLT we should be able to infer or predict their lack of success in later course work. This was not always the case. For example, as we examined the mark reporting statistics (the mark reporting statistics that we have access to, profile the final grades of each student who wrote the OSSLT) we found that in many cases students that passed the OSSLT ultimately did poorly in their course work. Likewise, we found that in many cases students that failed the OSSLT ultimately did well on their course

work. The following are just two of the many examples in support of the students that passed the OSSLT and ultimately did poorly in their coursework:

Student A

Grade 9 Applied English – 7%
 Grade 10 Applied English – 21%
 Grade Open Dramatic Arts – 1%
 Grade 9 Applied Foundations of Mathematics – 0%

Student B

Grade 12 Essentials English – 5%
 Grade 10 Open Civics – 35%
 Grade 10 Open Career Studies – 30%
 Grade 9 Applied Science – 11%
 Grade 12 Open Personal Fitness Activities – 23%

The above students scored poorly in their coursework and yet they passed the OSSLT. These students and the many others like them seriously challenge the predictive validity of the OSSLT.

The following are just two examples in support of the students that failed the OSSLT yet ultimately did well on their coursework:

Student S24

Grade 10 Applied English – 73
 Grade 10 Applied Science – 80
 Grade 10 Open Communications Technology – 75
 Grade 10 Open Learning Strategies – 85

Student S15

Grade 10 Open Visual Arts – 80
 Grade 10 Academic Spanish – 77
 Grade 10 Academic Principles of Mathematics – 67
 Grade 10 Academic Canadian History – 73

There were many more examples of students that failed the OSSLT yet ultimately did well on their course work. Our data

includes the marks and courses of all the students who failed the OSSLT. It becomes clear looking at these marks that there is little predictive validity between the OSSLT pass/fail and student course grades. If there was predictive validity then the course grades and the OSSLT pass/fail should be in agreement. Given the high-stakes nature of the test and that students' futures hinge on whether they earn an OSSD, this test must be 100% accurate. We cannot be satisfied with anything less. Looking at the marks in our data, it is unconscionable for us to claim that students who are literate enough to pass all of their course work are not literate enough to earn an OSSD based on one flawed test. In an article in *Our Schools/Our Selves*, Carlo Ricci (2004) has argued that both the test and the scoring are flawed.

It is unimaginable and absurd for any test or test developer to have the audacity to claim omniscience. This one narrow assessment indicator has the power to withhold a lifetime of opportunities. Literacy is never finished.

Criterion Validity

Strenio, Jr. (1981) states, "criterion validity asks if test scores match other possible indicators of student ability" (p. 91). This means that two instruments — for example, teachers and the OSSLT — should agree on who is literate enough to earn an OSSD. If there is a significant discrepancy between the teachers' judgments and the OSSLT's judgments then the test clearly lacks criterion validity. Therefore, the inferences we can make based on this test about students' literacy are not valid.

In order to test this we approached grade ten teachers and presented them with a sheet of instructions that included the following:

Please circle on the attendance sheet that has been provided those among your grade 10 students who you would consider NOT literate enough to earn an OSSD.

We chose grade ten teachers because the students who are interested in earning an OSSD are expected to write the OSSLT in grade ten. There were a total of 32 grade ten teachers at the school. Of the

32, we were able to survey 30. Of the 30, 13 were female and 17 were male. Nine teachers have been teaching for four years or less, 4 teachers have been teaching for between 5 and 10 years, and 17 teachers have been teaching for over 10 years. Only one of the 30 teachers had checked which of her/his students passed or failed the OSSLT and so she/he was aware of who among her/his students passed or failed; even with this information, the teacher still did not agree with the OSSLT results. She/he felt that one out of her/his 60 students was not literate enough to earn an OSSD. However, three of her/his 60 students failed the OSSLT and are therefore not eligible to earn an OSSD. It is also interesting to note that the student she/he identified is not among the students that failed the OSSLT. The rest of the teachers had not checked and so were not aware of who among of their students passed or failed the OSSLT.

We tabulated the number of teachers that say students should not earn an OSSD, the total number of teachers' judgments that are included in the study for this student, the number of instances where teachers and EQAO agree that the student should not earn an OSSD, and student course grades. For each of the 52 students we attempted to gather as many judgments per student as we could. That is why there were 52 students that have failed and 138 judgments. The students at this school have performed above the provincial average on the OSSLT.

The data above reveals the discrepancy between the teachers' judgments and that of the OSSLT. It lists all of the 52 students that failed the OSSLT at the school where we conducted our research. In almost all cases the teachers and the OSSLT disagreed on who they would consider not literate enough to earn an OSSD. In fact, in 138 separate judgments teachers identified only 16 students that they would consider not literate enough to earn an OSSD. Meanwhile, the OSSLT has identified 52 students that they consider not literate enough to earn an OSSD. In total, teachers and the OSSLT agreed on only 16 students. However, it is important to note that of the 16 agreements not all of the teachers teaching those students share the same opinion on whether they consider the respective students literate enough to earn an OSSD. For example, for student S21, one teacher and the OSSLT judged that the student is not literate enough

to earn an OSSD; however, four of this student's teachers judged that this student is literate enough to earn an OSSD. Not only do the OSSLT and teachers disagree, but teachers even disagree among themselves. This highlights the complexity involved in judging an individual to be literate or not.

Given this complexity we cannot rely on the inferences of one simple measure. And given the discrepancies between teacher's judgments and the OSSLT we have no choice but to stop making these invalid inferences. Having no confidence in the inferences made, we need to stop unfairly denying these students their earned right to an OSSD simply because they failed a flawed test.

Another frustrating illustration of the discrepancy between teachers' inferences and the OSSLT's inferences is highlighted by student S13. This student scored 90% in his grade 10 English applied course (students in Ontario are streamed either into academic or applied English courses), and despite this high impressive score has failed the OSSLT. How can a student score 90% in English and fail a literacy test? This becomes more shocking when we consider that the school in general spends a significant amount of time and energy teaching to the test (Ricci, Forthcoming), and that the English department in particular models many of its assignments on the format of the OSSLT; therefore, you would think that given the similarity between course work and the format of the OSSLT, that students who pass their course work will pass the OSSLT — this significant discrepancy and others like it underscore that this is not the case, and this problem cannot continue to be ignored.

We also organized the data by focusing on teachers. The data lists the number of students that each teacher says should not earn an OSSD; for each teacher, the number of students that the EQAO says should not earn an OSSD; the number of students where teacher and EQAO agree that student should not earn an OSSD; and for each teacher, the total number of students. Just like the data that was organized by focusing on students, this data reveals the discrepancy between the teachers' judgments and that of the OSSLT. For example, of T15's 48 students the OSSLT judged that 10 are not literate enough to earn an OSSD; yet, the teacher has identified that only one of his/her students is not literate enough to earn an OSSD.

Furthermore, they only agree on one student. This means that they disagree on whether 9 of the students that have failed the OSSLT are literate to earn an OSSD. Another interesting case is T21. This teacher judged that 6 of his/her students are not literate enough to earn an OSSD and the OSSLT judged that 5 students are not literate enough to earn an OSSD. The absurdity is that despite the high number of students that they have judged not literate enough to earn an OSSD, there has not been one instance of agreement.

Once again, given the high instances of disagreement between the OSSLT and teachers' judgments, testing must stop. It is a waste of time, money, human energy, and considering the invalid inferences made from it, it's an unjust waste of human potential as well.

Conclusion

This study has done more than to just show the discrepancies between teachers and the OSSLT; it highlights the complexity of judging students and the injustice of denying them an OSSD based on a simple measure of one moment in a student's life. In sum, the test is subject to construct under-representation by being too narrow and failing to consider other dimensions of literacy. It also is subject to construct-irrelevant variance by ignoring the difference between language and literacy. As well, since the OSSLT does not predict how well or poorly a student does on coursework it lacks predictive validity. Finally, given the nearly ubiquitous discrepancies between the OSSLT and the teachers' judgments, the test cannot be considered to have criterion validity. As mentioned earlier Ary et al. point out that validity is the most important consideration in developing and evaluating measuring instruments. A test's worth is measured by whether the inferences we can make based on the test are valid. Since we would argue that all standardized tests fall prey to problems with validity, can we afford to continue misleading students and the public; especially, given the high stakes nature of this test in particular and many standardized tests in general?

Dr. Carlo Ricci teaches in the Education Department at Nippissing University in North Bay. Steven Taylor is an English teacher at a secondary school.

BIBLIOGRAPHY

Ary, Donald & Jacobs, Lucy Chesser, & Razavieh, Asghar. (2002). *Introduction to research in education*. (6th edition). Belmont, CA: Wadsworth/Thomson Learning.

Helping students improve their literacy and math skills. (2003, April 24). Toronto, ON: Ontario Ministry of Education. Retrieved February 28, 2004, from the World Wide Web: <http://www.edu.gov.on.ca/eng/document/nr/03.03/bg0325.html>

The provincial secondary school test. (1999). Toronto, ON: Ontario Ministry of Education. Retrieved February 28, 2004, from the World Wide Web: <http://www.edu.gov.on.ca/eng/document/curricul/secondary/oss/oss.html#3.1.4>

Ricci, Carlo (Forthcoming). "The Case Against Standardized Testing and the Call for a Revitalization of Democracy." *The Review of Education, Pedagogy, and Cultural Studies*.

Ricci, Carlo. (2004, Winter). "Breaking the silence: a marker speaks out against standardized testing." *Our Schools/Our Selves*, 13 N.2 (#74), pp. 75-88.

Strenio, Andrew J., Jr. (1981). *The testing trap: How it can make or break your career and your children futures*. New York, New York: Rawson, Wade Publishers, Inc.