FRONTIER CENTRE FOR PUBLIC POLICY

POLICY SERIES

NO. 03

TO TEST OR NOT TO TEST?

THE CASE FOR STANDARDS TESTING IN MANITOBA'S PUBLIC SCHOOLS

JANUARY 2000

BY RODNEY A CLIFTON

ISSN 1491 – 7874

FRONTIER CENTRE FOR PUBLIC POLICY
POLICY SERIES

January 01, 2000 (03)

*To Test or Not to Test?

The Case for Standards Testing in Manitoba's Public Schools

The issue of standards testing in Manitoba's public schools has a complicated recent history. After years with no tests, the Department of Education had started to use them again in a limited way. But, they have become a contentious political issue and the Department is now beginning to retreat from its policy.

For the most part, standards provincial testing as a key component of evaluation in Manitoba's public schools disappeared more than 30 years ago. Only a remnant of the former regime remained; in the early part of the 1990's, Grade 12 students wrote province-wide exams in only one subject per year, and the marks they obtained counted for only 30% of their final grade .

Rodney Clifton *is a professor of Sociology of Education, at the Centre for Higher Education Research and Development at the Department of Educational Administration, Foundations, and Psychology, University of Manitoba. He has held academic positions at Memorial University in Newfoundland, the University of Stockholm in Sweden and the Australian Council for Educational Research, Melbourne, Australia. He has been extensively published in various academic and policy journals, including Policy Options, the Canadian Journal of Education, Sociology of Education, and the International Encyclopaedia of Education. Dr. Clifton is a native of Jasper, Alberta. He has a M.Ed. from the University of Alberta, a PhD from the University of Toronto and a PhD from the University of Stockholm.*

The Frontier Centre is an independent, non-profit organization founded to undertake research and education in support of economic growth and social outcomes which will enhance the quality of life in our communities.

Through a variety of publications and public forums, the Frontier Centre explores policy changes required to make the Prairies a winner in the new economy. It also provides new insights into solving important issues facing our cities, towns and provinces. These include improving the performance of government expenditures in important areas like local government, education, health and social policy.

A small, full-time professional staff and an advisory board of scholars research, write about and communicate new policy ideas, sharing them with the media, decision-makers and opinion leaders throughout the Prairies.

This work has been generously supported by the Canadian Donner Foundation and other individual and corporate supporters

FRONTIER CENTRE FOR PUBLIC POLICY
Suite 25 Lombard Concourse - One Lombard Place
Winnipeg, Manitoba Canada R3B 0X3
Tel: (204) 957-1567 Fax: (204) 957-1570 E-mail: newideas@fcpp.org

To Test or Not To Test?

The Case for Standards Testing in Manitoba's Public Schools

Background

This paper will examine the question of standards testing in public schools. What is its purpose? Are standards tests an effective and acceptable tool for measuring public school performance? Should we keep them or discard them?

The issue of standards testing in Manitoba's public schools has a complicated recent history. After years with no tests, the Department of Education had started to use them again in a limited way. But, they have become a contentious political issue and the Department is now beginning to retreat from its policy.

For the most part, standards provincial testing as a key component of evaluation in Manitoba's public schools disappeared more than 30 years ago. Only a remnant of the former regime remained; in the early part of the 1990's, Grade 12 students wrote province-wide exams in only one subject per year, and the marks they obtained counted for only 30% of their final grade .

As part of a project called the Student Achievement Indicators Program, in 1992 the Canadian Council of Ministers of Education initiated a national math test, written by 47,000 students, aged 13 and 16, selected from all provinces except Saskatchewan, which refused to participate. The random student samples were tested in April 1993, and the results released in December of that year.

The news was not good. A total of 5,900 Manitoba students took the test, designed to assess math knowledge and problem-solving skills. Although pupils in French immersion schools fared much better than students in public schools, the tests revealed that 40% of those in Grade 8 programs taught in English scored below their grade level and below the Canadian average. Sixteen-year-olds fared only slightly better. Manitoba students taught in English had the lowest test marks in Canada.

A year later, in April of 1994, the Council assessed language arts skills in the same manner, this time widening the national sample to 58,000 students. The 5,500 pupils in Manitoba schools who wrote the tests achieved a higher national ranking than was the case with the math evaluations, with scores slightly above the national average. In this instance, Manitoba's French immersion students pulled the province's ranking down, with similar scores in reading skills but much lower ones in writing skills.

In April of 1996, the Council continued its assessment, with a student sample of 37,500 writing tests in the subject of science. Once again, Manitoba pupils scored just below the national average.

In July of 1994, the Minister of Education at the time, Clayton Manness, announced plans to reintroduce standards testing in Manitoba's public school Manitoba's Department of Education developed the test packages over the next few years. The actual testing started in June, 1997 and was conducted in 1998 and 1999.

Introduction

The formal tests, called "standards tests," were designed for students in Grade 3, Grade 6, Senior 1 (Grade 9), and Senior 4 (Grade 12). These tests reflect the concern that the Ministry has for assessing the "essential learning" of students at critical points in their educational lives, when the students are about 9, 12, 15, and 18 years of age. The essential learning comprises, primarily, English Language Arts and Mathematics, and secondarily, French, biology, physics, etc. Obviously, it is impossible for anyone to predict the future with complete accuracy so the curriculum and testing committees have made a great number of assumptions in identifying the essential things that students need to know 5, 10, or 20 years from now.

In Renewing education: New directions: The action plan, the Ministry states: "Outcomes are concise descriptions of the knowledge and skills that students are expected to learn in a course or grade level in a subject area." "Standards are descriptions of the expected levels of student performance in relation to grade- and subject-specific outcomes." "They allow for the consistent assessment of student performance." In saying this, the Ministry, representing the concerns of many Manitobans, has come to realize that developing effective monitoring systems and using standards tests can be helpful in ensuring that effective teaching and effective administration of schools takes place.

Since the Action plan was published in January 1995 -- 5 years ago -- many have criticized the development and implementation of standards tests in Manitoba. Some critics have said that these tests are unreliable, some have said that teachers are being forced to teach to the tests, and others have said that the tests are too expensive (see MacPherson, 1994; Wiebe, 1999).

Late in August of 1999, the government responded to opposition to the testing by switching gears. It announced that compulsory testing would continue in Senior 4, but would become voluntary in Grades 3 and 6 and in Senior 1. School divisions would be allowed to test at the lower levels if they wished, but they would no longer be required to do so.

Many of the criticisms that inspired this policy reversal are misplaced. This paper answers six questions brought forward by the critics, questions that people, citizens, parents, and teachers, who are interested in, and open-minded about, standards tests may want answered. Obviously, there are many more questions that need to be answered, but this is a beginning.

What are standards tests?

If you review the literature on standards tests, you will not find this concept. Rather, you will find the concepts "standardized tests," "criterion-referenced tests," and "norm-referenced tests". Standards tests are "relatively" standardized criterion-referenced tests that have items derived, and weighted, in terms of the objectives of the curriculum.

To call a test "standardized" means that it is a relatively objective test that yields the same score for all students across the province that achieve the same performance outcome. To call a test "criterion-referenced" means that it measures student achievement against a predetermined standard, or criterion, such as a specific set of grade 6 reading skills. In other words, good standards tests cover the material that the students were expected to cover in the curriculum at the level that the committees of teachers and specialists, who developed the curriculum and designed the test, thought was adequate for the specific subject and the specific grade level.

In addition, proper procedures would have ensured that the items were pre-tested on samples of students and rewritten to eliminate ambiguities and to ensure that the test represents, as fully as possible, the curriculum. Moreover, the test is marked by a number of teachers, who are specifically trained and who follow a protocol, so that, to the best of their judgment, the same level of achievement from students in various schools and divisions receive the same score.

In the Action plan, an example is given for Grade 3 Mathematics. The objective of mathematics at the Grade 3 level is to have students use 2- and 3-digit whole numbers to answer questions in four areas: patterns and relations, statistics and probability, shape and space, and number concepts and operations. A sample of the items used in the 1997 standards test are shown in the appendix. This test represents a sample of items that students are expected to understand. If students meet the minimum competency in understanding the core material, they meet the

To cite a commonplace example, the police use a standardized criterion-referenced test to determine if people are drunk while driving. It is called a breathalyzer, and it determines, with a relatively narrow margin of error, the blood-alcohol content of drivers. As we all know, some years ago the police used another test, one that had a greater margin of error. They drew a line on the ground and asked drivers who they suspected of being drunk to "walk the line." People who "fell off" the line would be charged with drunken driving. This test was not standardized; the performance of individuals was affected by a number of extraneous factors, their nervousness, whether they had the flu or inner ear infections and, more importantly, by the subjective judgment of the police officer conducting the test.

Another common criterion-referenced test that many people are familiar with is the instrument used by optometrists to help them prescribe glasses. You can, if you want, walk into a Shoppers Drug Mart and try on reading glasses and when you find a pair that help you see better, you can buy them. Glasses were obtained

FRONTIER CENTRE FOR PUBLIC POLICY
Suite 25 Lombard Concourse - One Lombard Place
Winnipeg, Manitoba Canada R3B 0X3
Tel: (204) 957-1567 Fax: (204) 957-1570 E-mail: newideas@fcpp.org

provincial standard.

Do other professionals use standards tests?

To repeat, "standards tests" are standardized criterion-referenced tests. Thousands of tests like this are used in many different fields. In fact, the history of science, both physical and social, has often been written as the history of the standardization of measurement. In the physical sciences, we have learned to measure things like weight, distance, mass, and time, and for economists, concepts like value, marginal utility and Gross Domestic Product on standardized instruments. Otis Dudley Duncan describes this phenomenon in his valuable 1984 book, Notes on social measurement: Historical and critical.

this way before prescriptions became the norm. Most of us would probably agree that optometrists using a standardized instrument are more likely to do a better job of helping us obtain the appropriate lenses.

Standardized criterion-referenced tests are also used by Medical, Engineering, and Law Societies to determine if graduates from accredited university programs have achieved acceptable standards of performance in both knowledge and skill. These professions need an examination that is independent of those given by universities for three reasons: first, they do not trust universities; second, they need to protect citizens from those who are potentially incompetent; and third, they need to protect the profession from lawsuits.

Again, standards tests are relatively objective tests that yield the same score for people who have the same performance. Unstandardized tests, such as "walking the line," are more often than not scored differently by different people on different occasions. Consequently, unstandardized tests are inherently unfair to some of the people who have been assessed.

Do people support the use of standards tests in schools?

Most of us would rather have the police use breathalyzers to determine if we are inebriated and laser guns to measure the speed at which we are driving rather than having them use the subjective instruments they previously employed. We have a good reason to support these standardized criterion-referenced tests. The reason is, of course, that we expect equality of treatment.

Nevertheless, when it comes to standards tests used to assess the academic achievement of students, there are a number of vocal critics. In the Winnipeg Free Press (September 29, 1999), for example, Karen Wiebe wrote a critical letter saying, in part: "We do not support standards testing....What does standards testing show? Absolutely nothing more than how completely a teacher uses class time to teach to the exam and prepare the kids for the exam. These are not reliable assessments and evaluation methods." A few days later, on October 2nd, Jan Speelman, President of the Manitoba Teachers Society, which has consistently opposed standards tests, is reported to have said: "We want (Doer) to get rid of Grade 3 testing this year, immediately." (Martin, 1999, A5).

This is quite typical of the critics of standards tests in education. It is surprising, however, that we do not hear the same criticisms about breathalyzers, laser-guns, optometric instruments, the standard assessment of air craft pilots,

professional certification exams, and the thousands of other standardized criterion-referenced tests that are used in other professions.

More importantly, using standard procedures, a Gallup public opinion poll conducted in the middle 1980s showed that approximately 94 percent of citizens supported the use of standardized tests for assessing the achievement of students. In fact, about 69 percent of the respondents were strongly supportive (Roberts and Clifton, 1995, p. 289). In other words, the critics, even though they are often very vocal, do not seem to represent the concerns of most citizens. Far from being against standards tests, the general public, particularly parents and many teachers, are in favour of them.

Are standards tests adequate for judging the achievement of students?

Good classroom assessment begins with a teacher's own observations and measurement of what students are learning. This type of assessment is called "authentic assessment", in the jargon of the profession. Teachers spend between 20 and 30 percent of their time assessing the work of students. Unfortunately, many teachers never take courses in psychometrics, the theory and practice of developing and administering tests, and many never study ways of improving the reliability and validity of their assessment procedures. Consequently, many teacher-constructed tests have rather low reliability and validity coefficients, something that many teachers do not know. Many in the field believe that all teachers should know psychometric theory and they should be able to demonstrate ways of creating reliable and valid assessment instruments for their students. In fact, we should test

their competencies before we certify them to teach and evaluate children.

Standards tests have been created by committees of teachers, specialists in psychometrics, and specialists in the subject-area. The evidence demonstrates that the psychometric procedures for creating good tests have often been followed and the psychometric properties of the tests have been formally assessed. Consequently, standardized tests, in comparison with teacher-created tests, are generally fairer to students, particularly to disadvantaged students, because they more adequately cover the curriculum and they more adequately measure the varying performances of the students. In fact, of all the assessment instruments that have been developed in the social sciences, the best instruments are standardized criterion-referenced achievement tests. Well-designed achievement tests have much higher reliability and validity coefficients than tests that have been developed to measure other social and psychological characteristics of students.

In addition, after the tests have been given, the Department of Education publishes school, divisional, and provincial norms so that teachers can determine where their students are functioning in comparison with other students in the province. Information from standards tests can be helpful to teachers, students, principals, superintendents, and parents. But this information can never replace the formal and informal assessments that teachers make of students day-by-day and week-by-week. In other words, the results from standards tests can supplement teachers' assessments with additional information; they cannot replace the "authentic" assessments that teachers already use, even though, for many teachers, their understanding and use of testing procedures could be substantially improved.

What do standards tests offer us that we cannot get without them? In one word, they offer "comparability." Comparability in the context of the "big picture." It isn't very useful for teachers to compare their students' achievement with students one room down the hall and then to use that information to make decisions about instruction. It isn't very useful for parents to compare their children's performance with a few other children and then to make decisions about the effectiveness of the teachers and the schools. These comparisons are too limited. We need to back away from this type of comparison to understand the situation from a broader perspective. This is what standards tests enable us to do -- to back off a bit and to see the big picture.

Are standards tests useful for judging the performance of teachers and schools?

There are two purposes in using standards tests. They are useful in helping to identify the strengths and weakness in the learning of individual students. This is the most important role for these tests. But they are also useful in helping teachers, students, school administrators, parents, and other citizens evaluate the effectiveness of instructional programs. There are more problems with the second purpose than with the first.

Some school administrators and teachers are justifiably concerned that the average performance of students in a single classroom, during a single year, will be used to evaluate the effectiveness of a teacher or a principal. But the achievement of students is affected by many variables other than the actual teaching that has taken place in the classroom. A number of students, for example, may be ill with the flu and this may affect their performances on a test,

standard or otherwise. In some classrooms, there is more than a 100 percent turnover rate of students during a school year. When these thing happen, the assessment of the students does not represent the effectiveness of a specific teacher or a specific principal. Consequently, it is often inappropriate to generalize from an assessment of individual students to an assessment of classrooms and schools.

All of this simply means that the results of standards tests must be interpreted sensibly; without careful study and without qualifications they cannot be used to attack teachers and principals. Standards tests can provide a valuable measure of student achievement and a valuable measure of instructional effectiveness, but they cannot be used as the only measure to reveal the strengths and weakness of students or the strengths and weaknesses of instructional programs.

On their own, tests are incapable of harming students, teachers, or principals. The way in which the results of tests are misused is potentially harmful. Critics of standards tests often overlook this important distinction, preferring to target the instruments as if they were the

The answer to this questions depends on the value citizens place on having information about students and programs of instruction that is collected independently of the information amassed by teachers.

The standards tests in Manitoba cost about $15 million per year. At face value, this looks like a considerable amount of money -- about $15 for each person in the province or $70 for each of the 220,000 students in provincial schools. However, from the perspective of the total cost of education, about $1.5 billion, the cost of the standards tests represents about one percent. The most important question is: Do citizens and parents think it is worth the cost?

In light of this massive public investment of about 19% of the Manitoba budget, it might be claimed just as a matter of prudence that we ought to measure how effective these funds have been in achieving the goal of a sound basic education for children. Independent firms who do not work for the investment managers, for instance, constantly audit investment portfolios of this size in the financial services field. Standards tests represent a performance check on those who spend our

FRONTIER CENTRE FOR PUBLIC POLICY
Suite 25 Lombard Concourse - One Lombard Place
Winnipeg, Manitoba Canada R3B 0X3
Tel: (204) 957-1567 Fax: (204) 957-1570 E-mail: newideas@fcpp.org

culprits. It is irresponsible to blame all testing problems on tests, especially if the tests are well-designed, while absolving the people who interpret the results of tests -- parents, teachers, principals, superintendents, and trustees -- from all responsibility. In education, people interpret the results of tests, and they need to understand what tests can and cannot tell them. Standards tests can be helpful if they are well designed, if they are used properly, and if people do not interpret them in ways they were never meant to be interpreted.

Are standards tests worth the money?

school dollars. Spending one percent of the budget to obtain this information seems like a bargain.

Conclusion

Few people question the usefulness of standards tests in other professions. Not many people would argue that standardized accounting procedures should be discontinued or that the police should be forbidden from using breathalyzers and laser-guns. Not many people would argue that optometric instruments should not be used by optometrists. Properly used, well-designed standards tests can give teachers and parents feedback to determine whether students have attained the desired learning objectives

---

and whether or not they are ready for new ones. Properly interpreted, the results of good tests can inform citizens about the effectiveness of instructional programs. Without a doubt, the tests must be well designed, interpreted with caution, and the results must not be over-generalized.

Many critics charge that the emphasis on standards tests has lead to an epidemic of "teaching to the test," and other critics have charged that standards tests "kill the creativity" in teaching and learning. These criticisms are often over-blown. First, if the tests are derived from the objectives of the course, teaching to the test is, in fact, teaching to the objectives. Second, the tests are minimum competency tests: which means that they cover the core objectives and there is considerable opportunity for teachers and students to be creative in moving beyond those objectives. Teachers must, however, help students achieve the core objectives, and if they do not, this fact will become self-evident when the students write the test. This is to be expected and accepted.

Finally, standards tests are not designed to predict the future. When a person passes a driver's test, a standard instrument that measures both knowledge and skill, no one can say with certainty that the person will never speed, run a red light, or have a serious accident. Similarly, when a child scores well on a standards English Language Arts test, no one can say that the person will be good at reading and writing throughout his or her life. No one can claim that standards tests are flawless. They can claim, however, that standards tests are useful because the responses of students are assessed in a relatively objective manner on items designed to measure the core objectives of a course in a way that is consistent and fair for all students in the province.

Appendix

Sample Grade 3 math test questions
From Manitoba 1997 Provincial Standards Test

PART A

1. Write the number that comes just before 90. _____

2. Write the number **nine hundred two**. _____

3. This is a rectangular prism.

How many faces does it have? _____

4. Circle the unit you would use to measure the height of a table.

metre litre kilometre

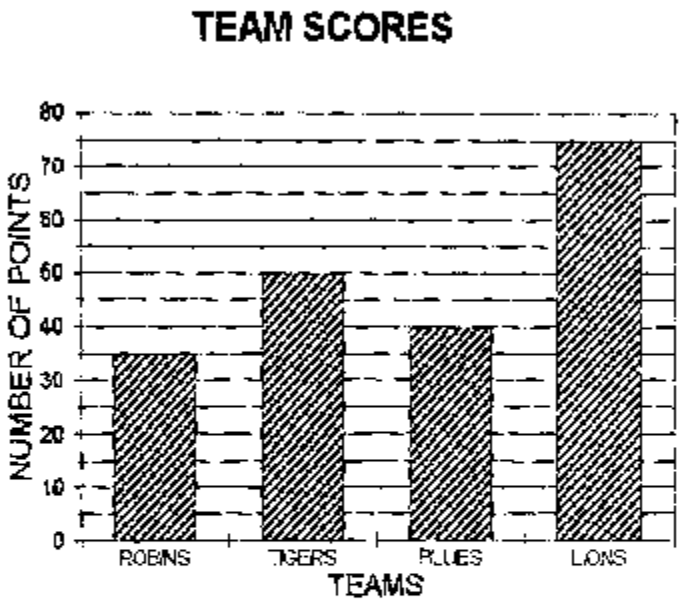
FRONTIER CENTRE FOR PUBLIC POLICY
Suite 25 Lombard Concourse - One Lombard Place
Winnipeg, Manitoba Canada R3B 0X3
Tel: (204) 957-1567 Fax: (204) 957-1570 E-mail: **newideas@fcpp.org**

9 Circle $\frac{1}{3}$ of this set of buttons.

Circle the shape that shows $\frac{1}{2}$ shaded.

FRONTIER CENTRE FOR PUBLIC POLICY
Suite 25 Lombard Concourse - One Lombard Place
Winnipeg, Manitoba Canada R3B 0X3
Tel: (204) 957-1567 Fax: (204) 957-1570 E-mail: newideas@fcpp.org

12 The children play games in teams.
This graph shows their scores.



How many points do the Robins score?

____ points

How many more points do the Lions score than the Tigers?

____ points

FRONTIER CENTRE FOR PUBLIC POLICY
Suite 25 Lombard Concourse - One Lombard Place
Winnipeg, Manitoba Canada R3B 0X3
Tel: (204) 957-1567 Fax: (204) 957-1570 E-mail: newideas@fcpp.org

Bibliography

APA (1985). Standards for educational and psychological testing. Washington, DC: American Psychological Association.

Duncan, Otis Dudley (1984). Notes on social measurement: Historical and critical. New York: Russell Sage Foundation.

Education Manitoba (1995). Renewing education: New directions: The action plan. Winnipeg, MB: Author.

Keeves, J. P. (1994) "Testing and the curriculum." In T. Husen and T.N. Postlethwaite (Eds.), The international encyclopedia of education, 2nd ed. Oxford: Pergamon Press.

Keeves, J. P. (1994) "Tests: Different types." In T. Husen and T.N. Postlethwaite (Eds.), The international encyclopedia of education, 2nd ed. Oxford: Pergamon Press.

Linn, R. L. (Ed.)(1989). Educational measurement, 3rd ed. New York: Macmillan.

MacPherson, Eric (1994). "Making sense of provincial testing." The Manitoba Teacher, 73,(3), 8.

Martin, Nick (October 2, 1999). "NDP hedge on the anti-Grade 3 testing promise." Winnipeg Free Press, A5.

Roberts, Lance W. and Clifton, Rodney A. (Eds.) (1995). Crosscurrents: Contemporary Canadian educational issues. Toronto, ON: Nelson Canada.

Rudman, H. C. (1977). "The standardized test flap: An effort to sort out fact from fiction, truth from deliberate hyperbole." Phi Delta Kappan, 79, 179-185.

Thorndike, R. L. (1982). Applied psychometrics. Boston, MA: Houghton Mifflin.

Walvoord, B. and Johnson-Anderson, V. (1998). Effective grading: A tool for learning and assessment. San Francisco, CA: Jossey-Bass.

Wiebe, Karen (September 29, 1999). "Standards testing useless exercise". Winnipeg Free Press, A13.

Worthern, B. R. and Spandel, V. (1991). "Putting the standardized test debate in perspective." Educational Leadership, February, 65-69.