

The Manitoba Grade Three Assessment:
A Generative Study of Teachers' Perceptions of the
Consequences and
Implications of Wide-Scale Formative Assessment on Classroom
Practice

by

Graham Bruce

A Thesis

Submitted to the Faculty of Graduate Studies

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF EDUCATION

Faculty of Educational Administration, Foundations, and

Psychology

University of Manitoba

Winnipeg, Manitoba

July 8, 2004

**THE UNIVERSITY OF MANITOBA
FACULTY OF GRADUATE STUDIES

COPYRIGHT PERMISSION**

**The Manitoba Grade Three Assessment:
A Generative Study of Teachers' Perceptions of the
Consequences and
Implications of Wide-Scale Formative Assessment on Classroom
Practice**

BY

Graham Bruce

**A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of
Manitoba in partial fulfillment of the requirement of the degree
Of
MASTER OF EDUCATION**

Graham Bruce © 2004

Permission has been granted to the Library of the University of Manitoba to lend or sell copies of this thesis/practicum, to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film, and to University Microfilms Inc. to publish an abstract of this thesis/practicum.

This reproduction or copy of this thesis has been made available by authority of the copyright owner solely for the purpose of private study and research, and may only be reproduced and copied as permitted by copyright laws or with express written authorization from the copyright owner.

Abstract

This study investigated teachers' perceptions of the consequences and implications of the formative Manitoba Grade Three Provincial Assessment on classroom practice. In 2000, this assessment replaced the year-end Grade Three Standards Test. The study examined teachers' knowledge of assessment practices, focusing on changes the mandated formative assessment has brought about in their daily lives.

Ten experienced Grade Three teachers from different English schools in an urban school division in Winnipeg were interviewed individually for sixty to ninety minutes about the former Grade Three Standards Test, the strengths and weaknesses of the current Grade Three Assessment, and the integration of the Grade Three Assessment into daily teaching and learning activities. A set of guiding questions provided a structure for the interview and teachers' comments were coded into categories and themes.

The results indicated that the majority of teachers tended to prefer the Assessment over the former Standards Test, especially in the area of Language Arts. They found it less stressful on both the students and themselves. The study found that the Assessment was integrated to a large extent into regular Language Arts activities but hardly at

all into regular classroom Math activities. Teachers also questioned the validity of the results because of a lack of standardization of tools and procedures. Professional judgment had replaced standardization and yet a wide range of difference in teacher assessment literacy was noted within the group. Although the Assessment intends to be a formative assessment, it is unclear if it succeeds fully, because it actually has certain features of a summative evaluation.

Acknowledgements

I would like to thank the members of my thesis committee Dr. Wayne Serebrin, Dr. John Stapleton, Dr. John Wiens, and especially Dr. Kelvin Seifert, my advisor, for their interest in and support and guidance of my endeavor.

I am indebted to the Grade Three teachers who took part in this study and shared with me their knowledge, perceptions, and joy of teaching.

I am also grateful to Heather Rosentreter for her word processing and proof-reading skills.

It is with great appreciation that I thank my wife Peggy-Anne for her constant love, support, and encouragement during the completion of this project.

Table of Contents

Abstract	2
Acknowledgements	4
Table of Contents	5
Chapter One: Assessment in Manitoba	
A Brief History of Province-Wide Assessment in Manitoba	7
Purposes and Practices of the "Formative" Grade Three Assessment	12
Statement of the Problem	19
Chapter Two: Published Literature Relevant to Provincial Assessments	
Standards Testing in Manitoba	21
Formative Assessment	25
Significance and Potential Value Research Findings.	27
Chapter Three: Methodology	34
Limitations of the Study	38
Chapter Four: Research Findings	
Setting the Stage: Who Were the Teachers?.....	41
Comparison of the Grade Three Assessment With the Grade Three Standards Test.....	42
Impact of the Grade Three Assessment on Student-Teacher Relationships.....	44
	49

Integration of the Grade Three Assessment into the Daily Life of the Classroom.....	55
Reviewing Concepts Prior to Assessing.....	63
Consistency of Results.....	65
Identifying Students Who Require Support.....	66
What Is Assessed and When It Is Assessed.....	68
Teacher Knowledge of Assessment.....	73
Chapter Five: Conclusions	77
References.....	86
Footnotes	91
Appendix A: Interview Guide	92
Appendix B: Literacy: Student Report	97
Appendix C: Numeracy: Student Report.....	98
Appendix D: Report to Division: Reading.....	99
Appendix E: Report to Division - Numeracy.....	100
Appendix F: Provincial Report 2002-2003.....	101

Chapter One: A Brief History of Province-Wide Assessment in Manitoba

Until the late 1960's, Grade Twelve students in Manitoba were required to write Departmental exams for every subject. The exam results counted for one hundred percent of their final grade. Following curriculum reviews at the system level, there was no province-wide student testing program in the 1970's and 1980's. In the early 1990's, Manitoba introduced mandatory Senior Four provincial examinations that counted for thirty percent of a student's final mark. Courses were tested on a rotational basis, Mathematics in 1991, Chemistry in 1994, and Geography and Social Studies in 1995.

In the past few years, the province of Manitoba has undergone many changes to province-wide assessment policies and practices. These policies and practices have had as their goal the improvement of educational achievement. In this regard, Manitoba is no different from governments across North America and around the world. Levin and Wiens (2003) note that the pressure to make changes in schools is "fueled by national and international test comparisons and driven by the belief that economic competitiveness rests increasingly on people's skills and knowledge (p.1).

Although education reform may vary from place to place, some

of the common elements across the English-speaking industrialized countries include:

greater specification of curriculum standards and outcomes...; more assessment of student outcomes and public reporting of the results on a school-by-school basis; greater opportunity for parents and students to choose the school the student attends; more pressure on teachers through measures that control their work, limit their pay, test their competence...; altered finance structures to reward schools that are able to increase academic results...; [and] within these policy frames, greater decentralization of managerial responsibility to individual schools (Levin and Wiens, pp.1-2).

The first of these recent changes occurred on July 4, 1994, when the Progressive Conservative Government of Manitoba presented a blueprint for the renewal of Kindergarten to Senior Four schools called *Renewing Education: New Directions*. Renewal would "provide students with enhanced opportunities to complete their schooling with the necessary skills and the levels of literacy and numeracy required to secure their futures and the well-being of Manitobans" (Manitoba Education Training and Youth, January 1995, p.3). This document laid the groundwork for a series

of new directions and actions in six interrelated priority areas. Levin and Wiens (May, 2003) note that some of the main changes made in education during the Conservative government's time in office were:

new curricula in many subject areas, with increased focus on literacy and mathematics; the introduction of Reading Recovery into elementary schools; an increase in provincial testing; an expansion of distance education; changes in teacher education; and the requirement that all schools have parent advisory councils (p.3).

In January 1995, a further document, *Renewing Education: New Directions, The Action Plan* described the setting of educational standards and policies related to student achievement. Standards of student achievement were developed for the compulsory core area subjects (Mathematics, English Language Arts, Science and Social Studies) and standards tests were planned for administration in those core subjects in Grade Three, Six, Senior One (Grade Nine), and Senior Four (Grade Twelve).

In May 1996, the first Mathematics standards test was administered to Grade Three students. This was followed in May 1998 with an additional standards test in English Language Arts. However, unlike the standards tests in Senior

One and Senior Four, the Grade Three test marks were not reported separately and therefore standards tests did not form a part of a student's final mark in Grade Three (Manitoba Education and Training, June 1995, p.22).

Dickens and Li (2002) note that there were five purposes of the standards-based testing program: first, to improve classroom instruction and student learning; second, to provide clear information about each student's skills and abilities in relation to the learning outcomes and performance standards; third, to provide schools, divisions and districts with benchmarks to review student learning; fourth, to contribute to a better understanding of student achievement across the province; and finally, to respond to strong public concerns for improved educational standards and enhanced accountability (Dickens and Li, 2002, p.1). These were noble purposes, but standards tests, and the Grade Three standards tests in particular, came in for strong opposition from both educators and parents. As part of their work at Manitoba Education, Training and Youth, Dickens and Li collated survey responses from teachers and parents. From these responses, there were four main criticisms of the Grade Three Standards Test: It was too late in the year for additional support and intervention for struggling students; two to three tests were too many for

eight-year old children; the standardized testing procedures were too rigorous and stressful for young children; and the public reporting of test results was of great concern (Dickens and Li, p.1). As well, any positive aspects were "greatly overshadowed by a climate of negativity about schools. Educators believed that schools and teachers were being criticized unfairly and being blamed for problems not of their making" (Levin and Wiens, 2003, pp.3-4).

In the fall of 1999, the New Democratic Party was elected in Manitoba. It quickly began work on an overall education plan that would focus on a small number of priorities that had support from educators, the public and research. The government was concerned with "trying to defuse the acrimony and bitterness that had developed and with building a climate of trust and common effort among educational partners" (Levin and Wiens, p.4). During its election campaign, the NDP promised to cancel all Grade Three standards testing, and the newly elected government followed through with its promise. In 2000, therefore, the standards tests that had been in place for only two or three years were therefore replaced with a province-wide, classroom-based, formative assessment in reading and numeracy. This mandated, wide-scale formative assessment was

a unique, radical departure from end of year standards testing.

The Purposes and Practices of the "Formative" Grade Three Assessment

Currently, as of 2004, the new Grade Three Assessment is administered during the first few months of the school year. Its primary purpose is to "provide meaningful information to parents regarding their child's foundation knowledge and skills in Reading and Numeracy¹ at Grade Three entry and at Grade Four entry for French Immersion Lecture" (Manitoba Education, Training and Youth, May 2002, p.2).

Furthermore:

The Grade Three Assessment is an approach to assessing foundation knowledge and skills of students. It focuses on a set of critical competencies in the areas of Reading, Lecture and Numeracy/Notions de calcul that are derived from the current provincial curriculum documents. The Grade Three Assessment is one component of an assessment process conducted as part of a teacher's regular daily classroom practice. Identified gaps in knowledge or skills become the basis for follow-up that might include changes to instruction to

¹ Numeracy skills are the ability to work with numbers in various ways and to solve math-related problems (Manitoba Education, Training and Youth, September 2001)

foster learning and/or further assessment of a diagnostic nature (Manitoba Education, Training and Youth, May 2002, p.2).

In the fall of each school year, Grade Three teachers gather assessment information on their students and/or select the most appropriate strategies to assess critical competencies from samples provided by the department, their school division, school and personal repertoires. An important source of assessment information is the student's Grade Two teacher.

Teachers then may begin the assessment with observations of all students. However, students performing consistently may demonstrate sufficient evidence of achievement in their daily work and further assessment may not be necessary.

Students performing inconsistently will require a more detailed systematic assessment. Some students may have been assessed recently and may only require an updated assessment. Additional assessments using the instruments identified by the department, or others, should occur where teachers feel they require more information to make an informed judgment (Manitoba Education, Training and Youth, May 2002, p.3).

It should be noted that this process outlined by Manitoba Education, Training and Youth is not the practice actually carried out by all school divisions. Some divisions mandate certain tools to be used to assess each competency and also mandate that every student will be assessed using a paper and pencil assessment.

The department expects teachers to use "professional judgment" when identifying the process for gathering information and does not expect that every student will go through a "comprehensive assessment on each competency" (Manitoba Education, Training and Youth, May 2002, p.3). In fact in many cases, teachers will be able to make an informed judgment about a student's level² on particular competencies, based on their daily work with their students (Manitoba Education, Training and Youth, May 2002, p.3).

Following the assessment, a report on all of the identified critical competencies in Literacy (Appendix B) and Numeracy (Appendix C) is sent home to parents by the end of the first reporting period of the school year. Forms are provided by the department for this. Schools must also compile and submit school-level results to their school division who in turn compile a divisional picture of the

² The three assessed levels of performance are: "needs on-going help;" "needs some help to meet expectations;" and "meets expectations."

assessment and submit it to the department (Appendices D & E). The compiled provincial results are presented in an annual report, which is posted on the department's web-site (Manitoba Education, Training and Youth, May 2002, p.6). Results from the 2002-2003 Provincial Report can be found in Appendix F.

It would seem that, on the surface, using a formative assessment at the beginning of the school year is a good thing for teachers to do. It provides them with information regarding each student and this information is then communicated to parents. However, the assessment is not without its problems. The assessment requires teachers to collect data on student performance in order to "base teaching decisions on solid data rather than on assumptions," so that "they can make adjustments early on to avoid the downward spiral of remediation" (Walter, Burger and Burger, 1995, p.39 cited in Schmoker, 1999, p.35). In fact teachers are expected to "avoid the downward spiral of remediation" by using the collected data to diagnose learning difficulties and plan a program of remediation. But this is not the only thing that the data are used for.

Manitoba Education and Youth collect and publish provincial Grade Three Assessment data which show the percentage of students at each level of performance in each

critical competency. In doing this, Manitoba Education and Youth would seem to be confusing a formative assessment with the "measurement-driven accountability" (Froese-Germain, p.18) of standards tests. Rather than using data to improve student performance, teachers, administrators and parents compare their child's performance to provincial averages. These provincial averages are compiled from data that comes from wide inconsistencies in where it is collected (an isolated northern community is probably quite different from an urban Winnipeg school), how it is collected, the tools used (some school divisions mandate tools to be used whereas other divisions leave it to the teacher's discretion), the classroom conditions, and experience level of the teachers. The formative assessment feels more like a summative evaluation. By being unclear as to the purpose of collecting the data and by using the data to measure and compare schools, Manitoba Education and Youth invite criticism that the assessment is used to hold teachers accountable. Leithwood has concerns that accountability policies do not work: some are unethical; some do not accomplish the intended purpose; they distract students from learning; they distract teachers from teaching; and, they are introduced for purposes other than teaching and learning (Leithwood, 2001, pp.2-5). Politicians have also used the published

provincial data to attempt to hold the government accountable for its policies:

Mr. Stuart Murray (Leader of the Official Opposition):
Mr. Speaker, can the Premier share with the House the results of the 2001 Grade 3 diagnostic assessments?...I would like to share with the First Minister that the results range from a high of 68 percent in terms of Grade 3 students meeting expectations when it comes to reading and interpreting graphs. However, on the other hand, only 42 percent on Grade 3 students are meeting expectation when it comes to recalling addition and subtraction to 10... Has the Premier identified the cause of why only 42 percent of Grade 3 students are meeting the Doer government's expectation to add and subtract to 10? (Oral Question Period in the Manitoba Legislature, May 14 2002).

Data then must be used "judiciously and with discretion" (Schmoker, 1999, p.35) so that consequences associated with publishing data do not go against the stated purposes and principles of the assessment.

The Grade Three Assessment directs teachers to use their professional judgment in terms of the amount of assessment each student requires and the methods and assessment tools used. But what if a teacher's professional

judgment is not very good? What if a teacher's professional judgment lacks experience, knowledge and a comfort level in sound classroom assessment practice? For sound classroom-based assessment to work well, a highly knowledgeable and skilled teaching force is required. Apart from a few professional development sessions mainly concentrating on how to interpret the results of the assessment, Manitoba Education and Youth has left the responsibility of staff development to school divisions and schools. As they are mandating a formative assessment based on strong classroom-based practice, perhaps Manitoba Education and Youth should take a stronger leadership role in staff development and should work closely with our universities to ensure that student teachers do not graduate without required courses in classroom assessment.

After the assessment is completed in the first few months of the school year, there is no follow-up by Manitoba Education and Youth to evaluate student progress and, in fact, there is no other assessment or testing required until Senior Four³, a gap of nine years. So the onus and responsibility for tracking and reporting on student progress rests almost entirely with the school. Would it be

³ Standards tests in Grade Six Language Arts and Senior One Mathematics are optional.

useful or appropriate for Manitoba Education and Youth to require a follow up report in June about students who needed "on-going help" in November to see if they were able to "meet expectations" on all competencies, or should this be left in teachers' hands?

Statement of the Problem

Levin and Wiens state that there is still work to do so that the Grade Three Assessment "becomes an integrated part of teachers' ongoing instruction as opposed to an externally mandated requirement that is unrelated to day-to-day teaching practices" (p.7). One cannot assume that because it is a mandated requirement, teachers have embraced the assessment and used it as a catalyst in changing their classroom practices. Because of that, many questions need to be answered. What changes in classroom practices have resulted from this mandated assessment? Have teachers increased their knowledge of sound assessment practices and to what extent are they implementing them in their classrooms? Has the content and methodology of their teaching had to change to accommodate the assessment? What effect has this assessment had on how teachers structure the beginning of their year and how has this affected their relationships with their students?

For all its good intentions, is mandated province-wide formative assessment actually working for teachers, students and parents? Can formative assessment be mandated and if so, is it making a difference in classrooms for teachers and students? It cannot be assumed that the assessment is implemented well or in the way that it was intended, or that teachers agree on its value. The best way to find out and get the stories behind the data is to talk to teachers about their beliefs, thoughts, fears, concerns and experiences with the Grade Three Assessment.

Chapter Two: Published Literature Relevant to Provincial Assessments

There is a great deal of research on summative assessments and standardized testing, and the debate as to the appropriateness and effectiveness of such testing has been ongoing for many years. The word *standardized* simply means that the same set of test questions is given the same way and under the exactly the same conditions to all students. Further, all tests are scored the same way (Chappuis and Chappuis, 2002, p.95). Centralized standardized testing is an entrenched part of the educational system in the United States costing billions of dollars annually and also spawning a multi-million dollar testing taking/preparation industry. In Canada, many provinces such as British Columbia, Alberta, Ontario and Quebec also have extensive year-end standardized testing in a wide variety of subjects. Alberta Learning tests all children in the core subjects in Grade Three, Six, and Nine to "ensure our education system maintains a high standard" (Alberta Learning, 2003, p.1).

Many other jurisdictions in North America also support the link between high standards and standardized testing and point to a good accountability system as "a tool that if properly sharpened and applied can be used by local school

leadership to improve student performance" (Johnson, Treisman and Fuller, 2000, p.20). "Similarly, when state-mandated standards and high-stakes tests are part of a larger system of accountability that includes local standards and local assessment, they are more likely to be helpful in holding schools accountable and in helping schools get better than is now the case." Thomas Sergiovanni (2000) goes on to say that used alone, "mandated standards and tests provide a dangerously narrow approach to accountability" (p.6).

One view, therefore, of how to reform and improve education is to test students vigorously and often and hold schools and school districts accountable for doing so. According to Levin (1998):

An emphasis on standards, accountability and testing has been a feature of reforms in many countries. Almost everywhere we find more large-scale testing of students and more reporting of the results of these tests than was the case a few years ago. Increasing national assessment is complemented by more and more international assessment and in both cases the results are used more overtly for public comparisons (p.133).

Manitoba departs from this trend in that it only requires its students to take two standards tests at the end of their

Grade Twelve (Senior Four) year. Manitoba also calls their criterion-referenced tests *Standards Tests* rather than Standardized Tests. Earlier testing in Grade Six and Grade Nine (Senior One) is optional for school divisions, but once a division decides to participate, teachers must administer the test to their students.

For all the funds that have been spent to test North America's children, there is growing dissent that testing is not producing the kinds of results that one might or should expect. "The U.S. private testing industry is a billion dollar a year business which, despite its growing influence on the education system, is virtually unregulated and unaccountable" (Froese-Germain, 1999, p.12). In both Canada and the U.S., the production of standardized tests is mainly in the hands of commercial publishers and non-school agencies. In Canada, standardized tests are developed and marketed by the major textbook publishers with profits going to the U.S. based parent companies (Barlow and Robertson, 1994, p.120). According to Barlow and Robertson (1994) "if testing created excellence, American students, the 'most-tested' students in the world, should be winning the educational sweepstakes" (p.117); instead, the U.S. consistently scores low on large-scale international assessments (Earl, 1995, p.5). In the 2000 Organization for

Economic Cooperation and Development (OECD) Program for International Assessment (PISA) study, fifteen year old American students placed 15th in the world in reading literacy, 18th in mathematics, and 14th in Science whereas students in other countries where such wide-scale testing is not practiced scored much higher (OECD, 2000, p.3-6).

James Popham (2001) believes that "today's high-stakes tests, as they are used in most settings, are doing serious harm to children" (p.1). He goes on to say that because of unsound, high-stakes testing programs, "many students are receiving educational experiences that are far less effective than they would have been if such programs had never been born" (p.1). He is not, however, opposed to educational testing and says that if properly built and sensibly used, educational tests can help teachers deliver effective instruction (p.15). He talks about the "unfilled potential" of using classroom tests solely to dish out students' grades and likens it to "using a laptop computer to keep your lap warm" (p.117).

Alfie Kohn (2000) is perhaps the most outspoken critic of standardized testing and believes that the more we learn about standardized testing, particularly in its high-stakes incarnation, the more likely we are to be appalled. "And the more we are appalled, the more inclined we will be to do

what is necessary to protect our children from this monster in the schools" (p.1). He feels that:

Standardized testing has swelled and mutated, like a creature in one of those old horror movies, to the point that it now threatens to swallow our schools whole. Of course, on the late, late show no one ever insists that the monster is really doing us a favor by making its victims more "accountable". In real life, plenty of people need to be convinced that these tests do not provide an objective measure of learning or a useful inducement to improve teaching, that they are not only unnecessary but highly dangerous (p.1).

Rick Stiggins (1996) also has serious concerns about the impact of standardized testing on students. He says:

An historical perspective reveals forty years of layer upon layer of ever more centralized testing costing billions of dollars with no discernible impact on the quality of our schools!...Obviously, the path to school improvement must lie in some other direction (Stiggins in Earl and Cousins, p.55).

Standards Testing in Manitoba

Some of the criticisms levelled at standardized tests, such as their construction by test businesses, multiple choice format and centralized scanned marking

do not apply to the Manitoba tests. Tests in Manitoba differ from those in other jurisdictions in important ways. Manitoba Education and Youth develop all standards tests in collaboration with teachers from across the province and the tests are marked locally, that is at the school or division level (Manitoba Education and Youth, June 2003, p.1). The format of Manitoba Education and Youth's tests has moved away from a one-time opportunity to take the test to tests that are now written over three or four days, or in the case of Senior Four Consumer Mathematics, written over a whole semester. The Grade Six English Language Arts test is written over a three day period (sixty minutes each day) and includes the activating activities of brainstorming, large-group discussion and group work. Students also have the freedom to choose a form for their writing task from a list of provided choices (Manitoba Education and Youth, June 2003, p.2).

The Senior Four English Language Arts test spans a four-day period and is also a far cry from the high-pressure, high-stakes format of many jurisdictions. The three hours of day one are spent activating and acquiring the theme of the test by jotting down ideas in response to prompts, viewing selections on the

theme, clarifying and extending understanding through discussion with others, and previewing the writing task. On day two, students plan their written test and identify their writing variables, manage ideas, and outline and draft their written text. On day three, students revise, edit and begin the final copy of their written text. Finally, on day four, students complete the final copy, proofread, and reflect on the test process (Manitoba Education and Youth, September 2002, p.2).

Formative Assessment

Rick Stiggins believes that the path to school improvement lies in a reevaluation centered not on *how we assess* student achievement, but on *how we use assessment* in the service of student success (italics in original, Stiggins, 1999, p.260) and in a direction away from standardized testing (Stiggins in Earl and Cousins, p.55). That direction is an increased awareness of the importance of sound classroom assessment practices that improve student performance and motivation, and a recognition that "the gathering and use of assessment information and insights must become a part of the ongoing learning process" (Shepard, 2001, p.1066). Black and William (1998b) note

that "assessment becomes 'formative assessment' when the evidence is actually used to adapt the teaching work to meet the needs" (p.2). Earl and Cousins (1995) add:

The most important assessment and evaluation activities already happen as a routine part of schooling; that teachers hold the key to its success and utility; and that they should retain that control if assessment and evaluation are to benefit the students (p.17).

The importance of the role of the classroom teacher is also supported by Stiggins (1999) who says:

If we seek excellence in education, then the time has come to invest whatever it takes to assure that every teacher is gathering dependable information about student learning, day to day and week to week, not just year to year. This action must be central to all future school improvement efforts, because *if assessment is not working effectively day to day in the classroom, then assessment at all other levels (district, state, national or international) represents a complete waste of time and money* (italics in original, p.263).

Therefore, the aim of assessment is primarily to educate and improve student performance, not merely to audit it (Wiggins, 1998, p.7). Wiggins uses the term *audit* to describe checking up on activities after they are over. He notes that many educators are coming to understand that assessment is of no value unless it is educative—that is, instructive to students, teachers, and school clients and overseers (Wiggins, 1998, p.7). Popham (2001) agrees with Wiggins that classroom assessment has a higher calling, namely to "improve the calibre of classroom instruction" (p.117). In fact, "assessment done properly should begin conversations about performance, not end them" (Wiggins, 1993, p.13).

The findings of two meta-analyses on assessment have had an impact on the thinking behind the Manitoba Grade Three Assessment. Terry Crooks' study in New Zealand on *The Impact of Classroom Evaluation Practices on Students* (1988), and Paul Black and Dylan Wiliam's (1998a) *Assessment and Classroom Learning* study in England each cover a decade of global research on the impact of assessment on students' learning. Black and Wiliam conducted a comprehensive meta-analysis and synthesis of over forty controlled studies on the impact of improved classroom assessment on subsequent student success as reflected in summative assessments. They

concluded that there are consistent and sizeable gains in standardized test scores directly attributable to prior differences in teacher's classroom assessment practices (1998b, p.3-4). The research indicates that improving learning through assessment depends on five factors:

- the provision of effective feedback to students—this should “be about the particular qualities of his or her work, with advice on what he or she can do to improve, and should avoid comparisons with other pupils” (1998b, p.9)
- the active involvement of students in their own learning—“opportunities for pupils to express their understanding should be designed into any piece of teaching” (1998b, p.11)
- adjusting teaching to take into account the results of assessment—“for assessment to function formatively, the results have to be used to adjust teaching and learning—so a significant aspect of any program will be the ways in which teachers do this” (1998b, p.5)
- a recognition of the profound influence assessment has on the motivation and self-esteem of students (both of which are crucial influences on learning)—whilst it can help all pupils, the formative

assessment gives "particularly good results with low achievers where it concentrates on specific problems with their work, and gives them both a clear understanding of what is wrong and achievable targets for putting it right" (1998b, p.9)

- the need for students to be able to assess themselves and understand how to improve—"self assessment by pupils, far from being a luxury, is in fact an essential component of formative assessment" and "pupils should be trained in self-assessment so that they can understand the main purposes of their learning and thereby grasp what they need to do to achieve" (1998b, p.10).

The framework for the design of the Grade Three Assessment in Manitoba took note of the work of Black and Wiliam and the Assessment Reform Group in England. According to Dickens and Li, there are six principles of the Grade Three Assessment (pp. 3-5). They are:

- assessment of student achievement is an integral part of teaching and learning
- the practice of assessment must incorporate a positive capacity building experience for teachers—this can be done through professional development

focusing on a complete understanding of how to use assessment as a teaching tool

- teachers need to be involved as the most important professionals in the assessment process
- using a multitude of methods to assess students is a sound practice
- parents are the best partners
- the search for alternative approaches to large-scale assessment.

The first three somewhat resemble Black and William's work, but the focus is a little different. Black and William's five factors are mainly focused on the student, and in particular how the student should be actively involved in his/her own learning and able to self assess and understand how to improve. Teachers should provide effective feedback to students with the focus again being on what the student can do to improve. Black and William speak of how assessment affects student self-esteem and motivation. Dickens and Li's first principle recognizes that the assessment of student achievement is an integral part of teaching and learning, but instead of focusing on the learner, Dickens and Li's principles mainly focus on the teacher. The teacher needs to be involved as the most important professional in the

assessment process and building the capacity of teachers through professional development is very important.

Dickens and Li have three principles that are different, yet complementary to Black and Wiliam's work. If students are to create their own understanding and relate new material to existing conceptions, "skills and knowledge should be presented, learned and assessed within the context of meaningful information, and supported by many opportunities to use the new material in a wide range of circumstances" (Earl and Cousins, 1995, p.9). This means that teachers provide multiple opportunities for students to demonstrate what they know and assess them in a variety of ways. Dickens and Li's fourth principle states that using a multitude of methods to assess students is a sound practice.

The fifth principle recognizes the important partnership role that parents have in the assessment process. Earl and Cousins say that parents are critical partners in the evaluation process who can become active partners in witnessing their child's learning. Parents can also have a powerful influence on how their children view themselves as learners (Earl and Cousins, 1995, p.19).

Following the Manitoba experience of Grade Three Standards Testing, and the increased focus in research on the value of sound classroom assessment practices and its

effects on teaching and learning, the last principle offers the idea of exploration "outside the box" of traditional large-scale testing. Dickens and Li say that the province should continue the search for alternative approaches to large-scale assessment. The formats of the existing Grade Six, Senior One and Senior Four Standards Test, and the Grade Three Assessment would seem to indicate that Manitoba Education and Youth has taken some major steps along the road of alternative large-scale assessment. The journey along this road is continuing. In May 2004, the provincial government announced that the optional standards tests currently being written in Grades Six and Senior One would be discontinued after June 2004. The Department is also currently working on a support document for teachers regarding classroom assessment. A consultation draft of this document is scheduled to be released in the fall of 2004.

Significance and Potential Value of the Research

The Grade Three Assessment has raised many questions as to the effectiveness of adapting formative assessment on a widespread scale and to using it as a solution to public concerns regarding equity and accountability. As noted earlier, it has given educators many things to ponder and ask questions about. Is in fact, formative assessment still perceived as summative evaluation in disguise that is simply

transferred to the beginning of the school year? Is mandated formative assessment actually useful for teaching? Does mandated formative assessment save time and effort, or cost more time and effort? Do teachers in general really consider the assessment an improvement?

To shed light on these broad, big picture questions, teachers were asked to share their perceptions of their experiences with the assessment. However, the research delved much deeper into the underlying factors behind the data. Are we, for instance, trying to make assumptions and educational meaning from flawed data? Would it not make more sense to go to the people who have collected the data and talk to them so that the underlying factors behind the data could be discussed, explored and analyzed? Some of the underlying factors could include the developmental level of the child, the time given to the assessment, how much review or re-teaching takes place prior to assessing, the quality of instruction, the knowledge and competency of teachers in classroom-based assessment, the effects of the summer interruption on a student's learning, and assessing Grade Two competencies at the beginning of the year instead of starting Grade Three work.

One way to explore these questions is to talk to the practitioners who use the tests, namely the teachers, about

their beliefs, thoughts, hopes, fears, concerns and experiences with the Grade Three Assessment. The teachers' comments might give clues about how Manitoba Education, Citizenship and Youth, school divisions, schools and educators could work together to revise and/or tailor current assessment procedures. Currently, Winnipeg School Division is an example of how differences in assessment beliefs and practices have led to an ongoing dispute.

The division is the largest school division in Manitoba and does not take part in the Provincial Grade Three Assessment. In its place, the division has implemented the Comprehensive Assessment Program (CAP) for students from Kindergarten to Grade Five. The CAP assesses students in the areas of Math, Language Arts, Social/Emotional, and Motor Skills, and is therefore more extensive than the Grade Three Assessment. Like the Grade Three Assessment, CAP takes place during the first two months of the school year and in November a report is sent to parents. Currently, the Winnipeg Teachers' Association (WTA) and the School Division are in dispute over this assessment. The WTA claims that the CAP is a "good idea gone bad" (WTA, 2004, p.1) and takes up far too much teaching time. WTA has published articles and appeared on local radio stations to insist that the CAP be reduced by 25-50%. The Division has maintained their right

to mandate the CAP and to require teachers to support it and refrain from voicing their opposition to it or face disciplinary action. This dispute is currently unresolved.

In addition, teacher comments could also suggest improvements in training teachers to use and interpret assessment data. Training could take place at University and/or in Professional Development sessions organized by school divisions.

Chapter Three: Methodology

In September 2003, Principals in an urban school division in Winnipeg were contacted by the author and asked to recommend teachers for the study who had five years or more teaching experience at the grade three level and also had experience with both the former Provincial Standards Test and the current Grade Three Assessment. From this recommended list, ten Grade Three teachers from ten different English schools teachers were selected based on the criteria and their willingness to participate. These teachers had knowledge and experience with the former Grade Three Standards Test as well as the current Grade Three Assessment and were able to speak about the changes they experienced between the year-end summative standards test and the beginning-of-year formative assessment.

Interviews took place from October to December, 2003. This qualitative study used in-depth interviews as the main source for its data base. Teachers were interviewed individually for sixty to ninety minutes. They were asked to talk about their experiences administering the Grade Three Assessment and the Grade Three Standards Test and the main differences between the two. A set of guiding questions provided structure to the interview, but there was also

flexibility that allowed areas of interest to the individual teacher to emerge. After a few initial questions that provided background information such as the number of years taught and the number of years at the current school, the interview questions were grouped into seven themes. Teachers were asked about their increased understanding and use of assessment, their integrated use of assessment, how they balanced the use of formative and summative assessment in their classrooms, any diverse pedagogical methodologies used, any changes in relationships between themselves and their students during the assessment, any changes in how and when curricular content was delivered, and if their knowledge of student performance had increased because of the Grade Three Assessment. Follow-up interviews and/or phone calls took place if more information or clarification of the initial interview was required.

A tape recorder was used to record the conversations and to create transcripts for later analysis. In addition to the tape transcripts, additional field notes from the interviews were used that noted such things as the body language of the participant, the atmosphere of the interview, and other background information that helped to provide a richer understanding of the participants' views.

Data were analyzed from January to March, 2004. The transcripts were read over a minimum of three times, each time with more focus. After the first reading, tentative or emerging themes were noted and categories and classifications of ideas and terms constructed. During subsequent readings, themes, concepts, category-schemes further emerged and were organized and coded. Excerpts were sorted by code and concepts and explicit hypotheses were formulated based on the categories, concepts and themes. Resulting data consisted of two hundred and seventy-eight pages of transcripts and notes. While the transcripts are therefore too large to include with this thesis, they are available on request from the author.

Data are portrayed as a written description of themes and concepts supported by specific examples and quotes from the transcripts. Where quotes have been used, they are referenced by interview number and transcript line number (e.g. Interview 3, lines 365-369). Some of the data lend themselves to the production of demographic tables.

Appropriate school division protocol was followed when contacting prospective subjects. To protect confidentiality, the names of each teacher and school were given a pseudonym and the school division was referred to only as an "urban school division in Winnipeg". The recorded tapes and all

field notes were stored securely at my home and were destroyed once the study was completed. My faculty advisor, Dr. Kelvin Seifert and I were the only persons who had access to the tapes and field notes. Permission to carry out the study was granted by the Education and Nursing Research Ethics Board (ENREB) of the University of Manitoba.

Limitations of the Study

This generative study examined the perceptions of ten Grade Three teachers in one school division in Winnipeg. It did not set out to be a comprehensive study of all Grade Three teachers in the city or the province, teachers from rural or remote northern jurisdictions, nor teachers of foreign language immersion programs.

A limitation of generative studies is that they surface issues for discussion and future study rather than provide answers and recommendations. In this study, teacher perceptions are presented in themes and some conclusions are reached based on those perceptions. The author has chosen to let the teacher perceptions speak for themselves, and has therefore, not used them to generate recommendations.

Teacher perceptions also have limitations too in that they are influenced by personal opinion and personal interest, and may in fact not be indicative of what is best for students or the school system.

Chapter Four: Findings

This chapter begins by summarizing some background information about the participating teachers. Following that, eight main themes have been used as chapter sub-headings as a way of organizing the data gathered from teacher interviews. These sub-headings examine the comparison between: the Grade Three Assessment and the Grade Three Standards Test, the impact of the Grade Three Assessment on student-teacher relationships, the integration of the Grade Three Assessment into the daily life of the classroom, the review of concepts prior to assessment, the consistency of the results, the identifying of students who require support, what is assessed and when it is assessed, and teacher knowledge of assessment in general.

Setting the Stage: Who Were the Teachers?

Figure 1 summarizes some background information about the participating teachers. The teachers who took part in this study were all very experienced in early-years education, although they had also taught a variety of grades from Kindergarten to Senior One. In addition, two individuals had extensive training and experience working as resource teachers. The average number of years that the teachers had taught was twenty-two and ranged from twelve to twenty-seven years. The average number of years that the

teachers had taught Grade Three was eight, and ranged from four to fourteen years. Class size varied from twenty per class to twenty-nine, with an average of twenty-four students. Eight of the teachers taught an entirely Grade Three level, but two taught a combined Grade Three/Four class.

Teacher Names*	Years teaching	Years teaching grade 3	Years at current school	Years administering Grade 3 Assessment	Class size	Class composition
Ross	24	9	6	4	21	3
Gavin	27	12	3	4	23	3
Christine	24	6	18	4	21	3
Cathy	17	4	15	4	22	3-4
Sally	23	7	10	4	29	3
Margaret	12	5	5	4	20	3
Annette	26	14	15	4	24	3-4
Joan	20	10	10	4	29	3
Lydia	26	6	7	4	22	3
Leanne	20	4	3	3	26	3

* pseudonyms

Figure 1: Demographics of teachers

The teachers had all been at their current schools for a minimum of three years and five teachers had been there for ten years or more. Two teachers had been at their current school for only three years because that is as long as the school had been opened. The Grade Three Assessment

was administered this year for the fourth time and nine of the teachers had administered the assessment since its inception. The other teacher in the study had administered the assessment for the last three years. In terms of the gender of the teachers, eight were female and two were male.

Comparison of the Grade Three Assessment with the former Grade Three Standards Test

Because the teachers had administered both the former Grade Three Standards Test and the current Grade Three Assessment, they were in a good position to compare the differences between the two. In general they felt that while the Standards Test and the Assessment were both time consuming, several felt that the newer Grade Three Assessment took *more* time to complete. Gavin said that both the Standards Test and the Assessment decreased the amount of teaching time he had available to him and as such he was not in favor of either of them (2, 254-256). Ross said that the Assessment eats into the first two months of the school year (1, 71-72), Christine said it was very time consuming especially in Math (3, 99-109), and Leanne and Sally both noted that it was extremely time consuming especially in the first year that it was implemented (10, 57, and 5, 63-63). In fact, Leanne felt that she had an additional half-time job just administering the assessment (10, 58-59). The

majority felt that the assessment currently took less time to complete than it had in the first year of inception, because familiarity and comfort with the requirements of the assessment led to streamlining their practices, as Sally said:

I find each year we just get better at it and it's easier. In fact, at lunchtime, the resource teacher for the Grade Three class asked me if we had done the Grade Three Assessment this year. I said yes and she asked why she hadn't heard about it. I said, because number one, we're less stressed and number two, we're so good! (5, 64-69).

Five teachers raised the issue of stress or pressure when describing the standards test. Ross found the test very stressful (1, 59), Annette said that it just "killed her trying to teach it all" (7, 133-134) and Lydia added that it was very stressful for students and teachers as they were under pressure to cover everything (9, 72-74). A further stressor for the students, according to Lydia, was that they could not get help from the teacher during the test (9, 82). Sally also spoke of the pressure of trying to get through the curriculum (5, 50-52). It is interesting to note that the assessment was not considered stressful by any of the teachers. This could perhaps be best summed up in the

comments of Joan who said, "It's much easier in the sense that it is just stuff that I would normally all do anyways right at the beginning of the year" (8, 82-84). However, Lydia cautioned that even though the assessment was more relaxed, it made teaching more fragmented and choppy at the beginning of the year (9, 85-89).

But some teachers also found the standards test very useful. These teachers noted, for example, that the test made sure that they had covered the curriculum and kept them on task (Ross, 1, 67), and focused all year on the skills that were to be tested (Margaret, 6, 41-42). The kids really knew their Math, said Annette, and this was probably due to the fact that she had done two to three times the amount of Math that she should have done! (7, 138-142). Ross and Joan also found the marking process to be a very beneficial learning experience in terms of their own professional development. Joan said that in her school the team examined the results from the test and made changes in their practices to develop identified weaknesses into strengths (8, 102-110). Five teachers said that the assessment should take place at the end of Grade Two because Grade Two competencies were what was being assessed. The teachers asked why not assess those competencies during or after they

have been taught rather than after a two to three month gap, more often than not with a different teacher?

In comparison to the Standards Test, the Assessment was seen as something that could assist teachers to focus on skills that would help programming for students. The information gleaned from the results of the assessment gave a good picture of where the kids were at (Annette, 7, 117-119) and could be used to adjust teaching. In addition, it was noted by Margaret that many of the divisional and department recommended assessments follow best practice and were therefore very good tools (6, 50-55). In fact they were the kind of tools that she would normally use, with or without the assessment. Joan also said that they were the same kind of things that she would normally use (8, 83). However, not everyone was in agreement about the tools themselves or how they were used. Annette was looking for more standardization of tools used if the results were to have any meaning (validity) (7, 120). She said that there was a great deal of difference between classrooms and schools in the tools used, their application, the time spent assessing each competency, and how results were interpreted (7, 124-127).

As well as the tools themselves, other teachers wondered if all of the things on the assessment actually

needed to be there. Ross and Joan felt that it was not important to assess all of the things currently being assessed at the beginning of the school year. Joan said that she would not collect all that information at once, but rather collect it immediately prior to teaching a particular unit (8, 88). Ross said that he would not teach some units of Math (such as measurement) until March to take into account the mental and physical development of his students (1, 75-77). It would have made more sense to him to assess measurement in March immediately prior to teaching it. In his opinion, assessing it in September was a waste of time because it would have to be done all over again in March (1, 555-557).

Teachers were asked if the current assessment was an improvement on the old Standards Test. Eight replied that they thought it was and one replied that it was not. One teacher said that neither was acceptable and that the only difference between the two was "a different political motivation" (2, 247). The group who thought the assessment was an improvement cited its ease of administration, the reduced stress on students and teachers, and their ability to use the information to program for students at the beginning of the year. However, Margaret said that she would really like to see the assessment at the end of Grade

One or the beginning of Grade Two, because the earlier student weaknesses are spotted the earlier supports could be put in place. She went on: "If the assessment is truly going to be an improvement over the Standards Test, the province needs to do a much better job of providing funding to put supports in place" (6, 202-204). Joan, although she felt that the assessment was an improvement on the Standards Test, conceded that the "Standards Test improved us as a learning community because we had to work as a Kindergarten to Grade Three team to improve our results" (8, 102-103). One teacher felt that the assessment was not an improvement over the old Standards Tests and cited the usefulness of the results, the professional development activity of marking together, focusing teachers on the curriculum, and the improvement of the Kindergarten to Grade Three team, as the main reasons for this (1, 62-69).

Impact of the Grade Three Assessment on Students

In many conversations that I have had with teachers over the last three or four years, the issue of how the assessment changed teacher student relationships at the beginning of the school year was one I frequently heard. Four years ago, when the assessment was first brought in, many teachers felt that the time that they had previously used for "getting to know" the students had been taken over

by assessment and instead of building nurturing relationships and making the students feel welcome, students were instead greeted with a barrage of tests. Teachers felt frustrated that this was not the best way to introduce children to Grade Three. In this study, I wanted to find out if my perceptions that the assessment had affected student teacher relationships were true and if things had changed in the four years that the assessment had been administered.

Most of the teachers did not feel that students were unduly stressed by the Grade Three Assessment, or did they feel that their relationships with students were affected between regular class time and assessment time. Their perceptions were probably due to the teachers' familiarity with the assessment and its administration, combined with their inevitable lack of knowledge of the student at the start of the year's first two months.

Teachers handled their lack of experience with the students in different ways. Five did not begin the assessment until after the second week of school, and used those two weeks to build relationships with their students. As Joan said, "I don't assess them for the first two weeks, but by then I've got a pretty good idea. I have been known to be wrong sometimes but..." (8, 357-360). Sally spent the first month developing socially responsible behavior with

her students using a program called Restitution and did not consider beginning the assessment until well into October (5, 96-97). Gavin began with the Language Arts assessment because the students were more comfortable with it and this helped him get to know them before beginning the more challenging Math assessment in October (2, 355-357). Margaret added that getting to know her students as learners was a lot more than knowing them as only a reader and a numerator and so for her this meant also knowing them socially and emotionally (6, 431-435). Cathy felt that the assessment actually helped her to get to know her students better (4,261-262) and Sally said that she never felt that the assessment hindered relationship building (5, 420-430). She felt that it only helped. It seemed as if the teachers worked the assessment into or around the things they felt were important at the beginning of the year (5,412-416).

Ross was the only teacher who expressed strong feelings that the assessment was not an appropriate way to start the year. He said:

I think that when they come in September, they're still quite little...some of them are only seven years old. They need time to settle into a routine, to get used to their new teacher, to get used to one another, to new rules, and to new expectations. Instead, we do all this

assessment with them and actually we spent most of the time the first six weeks assessing (1, 129-136).

Ross found this to be frustrating as his current class was a bit more immature than the last year and had a harder time settling into a routine. It took them the first two months to get into a routine, and he regretted that it was during those settling in months that the assessment took place (1,457-459).

It would seem then, that most teachers, managed to incorporate the assessment into their beginning of year activities quite well, at least for the Language Arts component. In Language Arts, most students were familiar with similar procedures used to assess their reading abilities on an on-going basis. Running records, reading surveys and story re-telling are a routine part of the regular assessment in the youngest grades in all of the schools, so completing a running record as part of a more formal assessment should cause little stress. Gavin referred to this as "business as usual" (2,392) and Annette said that the students "were not stressed at all" (7,202). Cathy noted however, that some of her students "got a little nervous because she was writing notes" (4, 460-461).

Margaret attributed students' stress to the students, not to the assessment: "Students who stereotypically stressed,

stressed, and students who were capable, confident readers didn't stress" (6, 522-5230. Rather than a stressful experience for students, in fact some teachers noted that the experience was an enjoyable one because the student and teacher were able to spend ten to fifteen minutes of uninterrupted time together (Ross 1, 431-432 and Christine). Christine said that the students really enjoyed the one on one because they love to read to her (3, 287-288). And so the possibility that the assessment had detrimentally changed relationships between teachers and students did not hold true for the Language Arts component of the assessment.

In the Math component, most teachers felt that there was little stress for the students but their responses were more mixed. Teachers like Joan and Ross "downplayed it, didn't make a big deal of it" (8, 443-444 and 1, 437) and Ross said "he was not even sure that the students knew they were doing it" (1, 439-440). Gavin said that the students had "blank stares at times, but they weren't scared or terrified" (2, 412-413). Cathy noted that the students just treated the assessment like an activity sheet (4, 468-469). Sally even said that the students "loved it because they knew that the teacher was going to help them" (5, 591-592). Four teachers however, were not so sure that the Math component did not create stress of some sort. Annette felt

that it was stress free for her students but caused her "a good deal of stress trying to get it done" (7, 378). Leanne said that for her students there was nervousness and stress during the timed math activity (10, 481-482). For Christine, story problems caused anxiety for her students who were struggling with the reading demands of the questions. Students who didn't appear to be stressed reading for a running record in Language Arts, felt stressed having to use their limited reading skills to try and solve a story problem in Math (3, 293-295). Two of the teachers, Margaret and Lydia said that the students were more stressed in the Math component because there are more competencies and more tasks involved and it look longer to do (6, 543, and 9, 86). (There are eight competencies to assess in Math compared to only three in Language Arts).

The difference in perception of the Language Arts and Math components may reflect the teachers' greater comfort in teaching Language Arts than Math. In speaking about Language Arts, all of the teachers appeared confident and relaxed and spoke of the assessment fitting in with what they would normally do in their classrooms. And as I already mentioned, assessment-like processes were already in place for Language Arts throughout the earlier grade levels, and students were familiar and comfortable with them. Reading activities were

therefore better integrated with the provincial assessments of it, Math competencies, on the other hand, tended to be assessed separately from units of study. In other words, a competency was assessed following only a little revision or re-teaching, and then this pattern was repeated until all of the eight competencies had been completed.

Integration of the Grade Three Assessment into the Daily Life of the Classroom

"The Grade Three Assessment is one component of an assessment process conducted as part of a teacher's regular daily classroom practice," (Manitoba Education, Training and Youth, 2002, p.2). The policy document goes on to say that "in many cases teachers will be able to make an informed judgment about a student's level on particular competencies, based on their daily work with students" (ibid, p.3). To identify the process for gathering assessment information, teachers are to use their professional judgment and the department does not "expect, nor consider it appropriate, for every student to go through a comprehensive assessment on each competency (ibid, p.3). For the last three years, the expectation of the department has been that teachers integrate the assessment into their classroom routines and do not submit every child to an assessment of each competency if it is clear, in the teacher's professional

judgment, that the child can clearly demonstrate that he/she is performing consistently.

But is this in fact what is happening in the classroom? How are teachers integrating the assessment into their regular activities and/or how many components of the assessment are carried out as stand alone activities? The policy document states that teachers may begin the assessment with observations of all students and that students performing consistently may demonstrate sufficient evidence of achievement in their daily work and will require little, or no further assessment (ibid, p.3). Are teachers content with observing that a student is performing consistently and therefore needs no further assessment, or, are teachers assessing all students on all competencies?

Teachers are, for the most part assessing all competencies with all students and are not integrating the assessment into their daily activities. Whereas in Language Arts reading surveys, story re-telling and running records are all a part of regular instruction, teachers still said that they give a separate assessment for the department report. For most of the teachers, the assessments in Math are not integrated at all but are stand-alone activities. Gavin said that it is very nice to advocate integrating the assessment into your regular daily teaching, but "it just

does not fly well to integrate it" (2, 203). Leanne tried as much as she could but simply couldn't "integrate every single topic in the math program in just a casual way into what [she was] doing and expect it was going fit" (10, 79-83). One reason put forward by Ross regarding the difficulty of integration was that because he had to report the competencies separately it made sense to him to assess them separately. He said: "They've divided them up so you have to assess them that way" (1, 203).

So why is it that there seems to be such a gulf between what the department expects of the teachers and what is actually happening in the classroom? The efficient use of time is one reason put forward by six teachers. This assessment is still seen as something mandated, but the sooner it can be completed and got out of the way, the sooner the teacher can get back to teaching the Grade Three curriculum rather than assessing Grade Two outcomes. One way to be efficient, especially in Math, is to assess each competency with all students at the same time—as a whole class test—instead of planning and preparing integrated assessment activities, and/or trying to use individualized assessments. Joan said:

I'd rather spend the time on review and try and stick with one assessment. Now, if I got something really

weird on the assessment, where I thought this kid misunderstood the question... where I'm surprised at what I got... then I would go back with that individual and sit down with them one-on-one and try to communicate and see whether they really don't get it or what's going on there (8, 397-403).

In contrast to Joan, Christine said that most of her assessments were one-on-one because her students required that kind of support and each assessment took a long time with each child. She also noted that, "if you are going to assess the child, you want to make sure that what you are saying is actually true, and that takes a lot of one-on-one" (3, 149-151). She still assessed the whole class together in such things as addition and subtraction facts because "to assess each student individually would have taken forever" (3, 108). It is interesting to note that Christine had one of the smallest classes—21—of the teachers studied. When assessing Language Arts, other teachers create one-on-one time by using teaching assistants and student teachers. They further use time when the rest of the class was otherwise occupied, such as during computer and library times.

After the whole-class assessments, four teachers said that they would follow up with individual student interviews if confirmation of results were needed. Two teachers noted

that having the whole class all do the same assessment at the same time also depended upon the make-up of their class. In some years, it was not possible to assess every student this way, and so they did more individual assessment. Six teachers said that they do differentiate the assessment in some way for students who normally require adaptations, such as those with autism and behaviour disorders. Adaptations included reading the assessment to the student, assisting with writing, allowing more time, and providing more one-on-one assistance.

Ross also felt that there was a greater degree of consistency and continuity if all students were assessed using the same tools—meaning that they should be tested as a group. Group testing was in fact what took place in his school, and Ross felt it was especially important if a school has more than one Grade Three class, in order to insure validity in the results (2, 190). Ross also was concerned that teachers in other divisions were assessing in idiosyncratic ways (using different tools), and he wondered whether “if ten of his students were not meeting expectations would this mean the same as ten students in another school who were also not meeting the expectation?” (1, 391-394).

Gavin said that it would be very difficult for parents in this community to accept that some students were being formally assessed and others not. By "formal assessment", he meant that completing a paper-and-pencil task instead of the teacher just observing them perform. For parents in his community, it was also an issue of fairness.

The expectation on the part of the parents in this community is that if their child is being assessed more formally than another child then they are concerned, they are very anxious that their child isn't meeting something and by contrast, the child who is doing well, the parents want a formal assessment to verify that yes indeed their child meets or exceeds the start of Grade Three. And so in this community, the expectation, and there are two Grade Three classrooms in this school, is that all students will be given the assessment in a very similar fashion (2, 181-190).

Fairness was also a factor in Leanne's community. She said that parents were very aware of the assessment and were asking their child if he or she had been assessed and so "a little anxiety comes from home sometimes too" (10, 502-505).

The Math section assesses eight competencies, all of which have to be completed in time for a parent report in November. For some of the teachers, certain topics (e.g.

measurement) would not be taught until later in the school year when the students were more mature cognitively and therefore more likely to be successful. As Margaret said: "Some parts of the assessment I wouldn't normally do, but it has to be done and so I set it up in a way that's the least intrusive" (6, 168-171). To devote time assessing a topic when it was not going to be taught was not considered a good use of time. Teachers also felt that they would have to assess all over again when they finally did begin teaching measurement. Under the circumstances, therefore, assessing the whole class, at the same time, using the same test was considered to be the best option.

Given the choice of observing and judging an individual student perform or administering a paper and pencil task to assess the same thing, almost all of the teachers in the study used the latter. Why is this so? Teachers felt that they needed hard evidence, something concrete, to support their observations of student performance; merely telling parents that they had seen their child perform adequately was not sufficient. Teachers spoke of parents wanting "proof" that their child was performing well. Annette said: "I want to make sure I've got exactly the information that I need to give to the parents and the department so that I really have a good picture and it is a really true picture"

(7, 181-184). Joan too noted that "there are some kids that I have no doubt that they can and will do it. But I still feel I need to do that" [assess every student in every competency] (8, 219-221). Since the proof teachers sought was in the form of written samples of work, they assessed nearly all of the Math competencies using paper and pencil tasks. A further reason for using paper and pencil tasks was that teachers felt that had to be able to support their evaluation on the parent report card. If their signature on the report indicated that a student was performing at a certain level, then they were accountable for that statement and needed to be able to back it up. As Lydia said:

I do like to have something written down as proof that the student can or cannot do it. Because on the evaluation sheet we still have to sign our names, and if the parents come back and question us, I like to be able to say that this is why I gave the student such and such (9, 163-168).

Why teachers could not simply use the Grade Two end of year report as evidence of competence was highlighted by Leanne:

The Grade Two report card doesn't tell me if the students can measure within two centimeters, if they can estimate within two centimeters, it simply has a score for measurement. So I felt, it put me in an

uncomfortable position. If I'm expected to truly say this and know it and prove it, I had to do it with every single student. If I was willing to let that go and say, oh yeah, their teachers gave them the top level in everything in math and so I'm not going to test this kid, then I don't think that would be a professional way of doing it quite frankly (10, 243-252).

Teachers know that the results from their class were examined at the school level and forwarded to their division, where they were compiled, and sent as divisional statistics to the department. The data were then compiled into a published provincial report that indicated how students in the province were performing. Because of this, the assessment took on a more significant role than a mere classroom-based formative assessment. Why this was so, the distinctions between formative and summative assessment, and between assessment and evaluation, will be examined in more depth in Chapter 5.

Reviewing Concepts Prior to Assessing

Teachers reported a wide range of practices about revision or re-teaching of concepts prior to assessment, ranging from Margaret who did almost nothing, to Leanne who taught, then assessed, re-taught, and assessed repeatedly

until her students could all perform competently (10, 153-156). Margaret did not review concepts with her students because it "would make the assessment unfair" (6, 115-116). She would "ask them to think about what they had done in Grade One and Grade Two, the tools they had used, and then do the activity prescribed as well as possible" (6, 120-123). Teachers who extensively reviewed concepts were in Margaret's view "cueing the learner for the test" (6, 131), an activity that she considered more a part of the former Standards Test than the current Assessment. Leanne, on the other hand, said that sometimes she merely had to jog the memory of students. For her, the main point about the assessment was:

To make sure they're where they need to be. So if I can bring them there, even if it means I'm helping them to do better than they would have if I hadn't...I do it because I'm doing it for the benefit of my students so that they know what they need to know and can feel good about themselves (10, 210-218).

Apart from Margaret, all teachers said that they spent more time reviewing and re-teaching Math concepts than Language Arts. Since some of the Math concepts had not been taught since December of Grade Two, teachers felt that it was fair to review what had been taught before assessing them. As

noted earlier, some teachers completed the assessment as soon as possible, whereas others administered four or five separate assessments, assisting the student to improve in between each one. There was little consistency among teachers in how much concept review took place prior to assessment, or how often the students were assessed on a given competency.

Consistency of Results

In this school division, teachers have been provided with a binder of assessment tools. These tools are a compilation of materials provided by the department and materials recommended by divisional consultants. Many of the teachers interviewed supplemented the assessment tools provided with tools of their own that they had used over the years. The division does not mandate that certain tools are to be used and so teachers are free to use the tools of their choice. Often, in schools where there is more than one Grade Three class, teachers agree to use the same tools because they feel that this gives more consistency to their results. But across the division different tools are being used to assess the same competency. As mentioned earlier, there is also a wide variety of practice relating to how much reviewing take place prior to assessing each competency. And there are also differences in how many times

each competency is assessed. Students in one school had one shot at each competency whereas students in another had five and six different assessments and chances to show that they were competent. Because of all these differences, the question that needs to be asked is just how valid and meaningful are the results that are published by the department. I would suspect that the large differences in procedures within this urban school division would be amplified when applied in the provincial context.

Identifying Students Who Require Support

The Grade Three Assessment Policy document states that "schools consult with parents if the assessment results suggest a need for additional support" (METY, 2002, p. 4). These additional supports could include "intervention strategies at school or at home and possible further assessment of a diagnostic nature" and "implementation of an education plan for student learning" (ibid, p. 4). The policy also asks teachers to consult with parents as soon as possible once needs are identified. But how many children, identified by the assessment as needing additional support, were previously unknown to the teacher? All of the teachers who were interviewed said none. In addition, how many of the students identified as needing additional support actually received additional support following the

assessment? Again, all of the teachers indicated that no additional supports were put in place. According to the teachers, the students who needed support had already been identified at their school without using the Assessment. For these students, appropriate actions had already been taken, including Individualized Education Plans in some cases. As Sally said:

The Student Services are wonderful at my school. I find that we're doing precision reading, we're doing guided reading...because of the programs we're running in the building, I don't think the kids are falling through the cracks like they used to (5, 640-644).

Not one of the interviewed teachers indicated that a new or adapted education plan had been put in place for a student as a result of the assessment. In the first few years of the assessment, teachers expected that once a student was identified as requiring additional support, funds would be provided for more teaching assistant time or specialist time. Experience showed, however, that this was not the case. Teachers learned that there was no additional funding following the assessment, and that the resources and supports they had before the assessment essentially remained the same afterwards.

What Is Assessed and When It Is Assessed

Grade Three students are assessed on eight critical competencies in Math and three in Language Arts. The competencies are skills that students are supposed to demonstrate by the end of Grade Two, and as such are Grade Two outcomes. For each competency, teachers rate a student's level of performance, either *needs on-going help*, *needs some help to meet expectations*, or *meets expectations*. In Language Arts, the critical competencies are all related to reading and are:

- Reflection—a student's ability to think about his or her own learning as a reader
- Oral Reading Skills and Strategies—a student's ability to use a variety of strategies to read, and
- Reading Comprehension—a student's ability to understand and draw conclusions from text.

Most of the teachers said that they were comfortable with what was being assessed and when it was assessed in Language Arts. They noted that they were familiar with running records, comprehension and re-telling activities. In fact several said that they would carry out similar assessments even if the provincial assessment was not in place. They needed to know for themselves the reading abilities of their

students. Students appreciated the opportunity to read one-on-one with the teacher, even when the activity was part of the Assessment activities. Teachers took advantage of library time, computer time, time when the Instructional Assistant was with the class, and time that a student teacher was with the class to complete the assessment. Because the assessment mirrored current classroom practice, it was seen as far less of an imposition or intrusion.

There were some less positive opinions, however, about the eight critical competencies assessed in Math:

- Students sorts objects using one mathematical attribute; identifies attributes such as shape and size
- Student selects the appropriate standard unit; estimates and measures length
- Student recalls addition facts to 10; and, student recalls subtraction facts to 10
- Student represents and compares numbers, using terms such as even, odd, more, less, same as, to 100
- Student understands place value to 100
- Student identifies, extends, and describes mathematical (repeating and growing) patterns
- Student solves and creates addition and subtraction story problems, and

- Student reads and interprets graphs.

Teachers identified three main areas of concern about this part of the assessment. First, because the Math concerned Grade Two outcomes, some teachers felt it would have been better to do it at the end of Grade Two. At this time, students would not have had time to forget what had been taught. It would have been fresh in their minds. Cathy said what she thought she was really doing was assessing what the students had forgotten over the summer (4,204). However, I am not sure that this was the case as Leanne noted that some of the topics had not been taught since December of Grade Two and so they would not be fresh in the minds of the students even at the end of Grade Two (10,286). However, those topics would have been assessed by the teacher that had actually taught them.

Second, in doing the Math Assessment during the first two months of the year, teachers felt unable to teach the Grade Three curriculum. Some of the teachers (for example, Annette) said that they were not able to begin the Grade Three curriculum and the delay created stress. Annette, who was also teaching a Grade Three/Four combined class for the first time, put it like this: "I thought, I have hardly taught any Grade Three Math at all. I wasn't really doing Grade Three outcomes at all. It was basically Grade Two and

Four" (6,484-486). For Leanne, it was a matter of accepting that the assessment "became her Grade Three curriculum for a couple of months" and she had to stop beating herself up about it (10,636-638). Other teachers, however, were able to integrate assessing Grade Two outcomes with the Grade Three curriculum. Margaret said that it was a matter of spiraling the curriculum, teachers are always "building on the knowledge that the kids have, and once you leave a topic you need to return to it at some time in the future and build from where you left off" (6,460-462). She felt that the divide between Grade Two and Three curriculum was somewhat artificial and could be easily handled by a skilled and well organized teacher (6,684-690).

Third, teachers were concerned that topics on the Math Assessment happened in isolation from when the topics were normally taught in class. They felt that they were teaching or reviewing a topic, assessing it, and then covering it over again—eight times. It seemed like a shallow approach and not how they would teach normally. Sally felt that the Assessment caused her to cover topics "an inch deep and a mile wide" rather than "inch wide and a mile deep" (5,502). For Joan, doing a formative assessment removed in time from when you teach that topic meant that she had to do another assessment just prior to beginning teaching the topic. The

students' skills and abilities would have changed in the intervening time, and so the initial assessment results would be invalid and of limited use (8, 485-489). For other teachers, assessing all of the math competencies at the beginning of the year was unfair and demotivating. If a student did not do well in many of the competencies, his/her self esteem would fall and he would have a provincial report that saying that he was not doing well. As Lydia said:

"It's a little unfair to just throw, you know, do you remember all the names of the 3-D shapes and can you do patterning and do you know your facts and can you subtract and add and all of these other things all in one blow" (9, 150-154).

Assessment in isolation from when teaching takes place is at odds with Black and William's ideas of improving student performance through formative assessment (1998b, 9-11). In such a case, can teachers adjust their teaching to take into account the results of this assessment? Likewise, one of the six principles of the Grade Three assessment outlined by Dickens and Li is that student assessment is an integral part of teaching and learning (2002,3-5. Having an assessment at a time other than when the topic is taught would not seem to allow teachers to adjust their teaching to take into account the assessment results, nor would the

assessment be an integral part of teaching and learning. Cathy highlighted the dilemma: "let's stagger [the assessment] maybe throughout the year where we can... but even then staggering it doesn't make sense because you're not going to start assessing something you had done in Grade Two in January or March" (4,514-516). This raises a question: Does the assessment need to change to allow teachers to assess each competency immediately prior to teaching it; or, do classroom practices need to change to meet the demands of the assessment? Perhaps part of the answer can be found in this next section that looks at the teachers' knowledge of assessment.

Teacher Knowledge of Assessment

In preparation for the implementation of the Grade Three Assessment in the fall of 2000, the division had provided professional development workshops for teachers. In June 2000, the Director of Curriculum and the divisional team of consultants and teachers had taken staff through the Assessment Policy outlining the changes from the previous Standards Tests. A binder of materials (prepared by a team of Grade Three teachers) had also been provided to each teacher consisting of department produced and divisionally produced assessment tools. Teachers were free to use these tools, or their own tools to assess the students.

During the first two years of the assessment, Grade Three teachers were brought together three times to provide evaluative feedback on the new Assessment, and its implementation. Their comments were forwarded to the Department of Education, where many of the points were incorporated into the following year's Assessment. Because of this experience, teachers, on the whole, felt listened to, and cooperated with the Assessment more fully.

An example of how teachers influenced the assessment related to the Math competency "*student recalls addition and subtraction facts to 10.*" Teachers had indicated that there was a wide discrepancy in the results between addition and subtraction. In fact, student skills in addition were always much better than their skills in subtraction, and therefore combining the two scores was not accurate nor very meaningful. The teachers were pleased that their comments were listened to and that in the following year's assessment, this competency was divided into addition and subtraction and reported on separately. Teachers also had concerns that too many competencies had been assessed on the original revision of the Assessment. Again, however, their concerns were heard, and the number of competencies on the math assessment was reduced to eight in subsequent Assessments.

In terms of professional development, nine teachers said that they felt well supported by their division. They felt that they had received sufficient professional development, especially in the first two years of the assessment, to understand and administer the assessment. However, Lydia noted that there had not been anything organized by the division in the last couple of years (9, 244). When asked if he felt supported by divisional professional development, Ross said that he did, but he also felt that it was his own professional responsibility to take charge of his own learning and make sure that "he knew what he needed to know" (1, 348-349). In terms of whether or not the assessment itself had improved knowledge of assessment practices, the teachers were divided, with some saying that the assessment had helped and others saying that it had not. As experienced teachers, knowledge of assessment had accumulated over the years by reading, attending conferences and divisional in-services, school in-services, and, most importantly, through discussions with colleagues. None of the study group felt that university courses had given them a good background in assessment, although, as one teacher laughingly pointed out, university had been a long time ago.

Even though all teachers had received the same professional development about assessment, and even though

teachers said that they felt well prepared, the interviews suggested important differences among them in knowledge of assessment practices. Some teachers were able to clearly articulate the principles of, and differences between formative and summative assessment, and how each was used in their classroom. An example of this was Margaret, who put forward the idea of dividing formative assessment into three categories:

There's observational formative assessment, things like collecting anecdotal observations. There is product or process formative assessment where students are creating products or they are going through a process that you see, and then there's conversational assessment where you're conferencing with students, and with parents (6, 319-325).

On the other hand, some teachers were unable to define formative and summative assessment, and/or confused the two terms, and were unable to say how they used each in their classroom.

Teachers said that, on the whole, they did more formative assessment than summative assessment with their students. Answers ranged from "I do much more formative than summative" to the more specific "about seventy percent of my assessment is formative and thirty percent summative".

Chapter Five: Conclusions

The Grade Three Assessment was implemented in the fall of 2000 in response to parents', teachers', and school administrators' concerns that the Standards Test was not appropriate for Grade Three children. Eight year old students were thought to be too young to undergo the eight days of testing in language arts and math. The Assessment was considered stressful for both students and teachers and teachers complained of spending inordinate amounts of time preparing students in Language Arts and Math to the detriment of the rest of the curriculum. The Standards Tests were also costly to prepare and mark. The newly elected NDP government made good on their election promise to abolish the Standards Test and replace it with something that they hoped would better meet the needs of students, parents, and teachers. That "something" was the Grade Three Assessment.

All of the teachers interviewed felt that the Assessment created very little stress for students, and felt that the stress for teachers was manageable. Now in its fourth year, the demands of assessment time, result collating, and report writing are very real, but are being handled well by this experienced group of teachers. The

resentment about having to do a mandated assessment that was noticeable in the fall of 2000, has now reduced, and been replaced by a "dull resignation" among the teachers. In this sense, the Department of Education has successfully shifted from an end-of-year standards test to a beginning-of-year assessment.

In the current literature, the terms formative and summative assessment are now falling out of favor and being replaced with assessment for learning (formative) and assessment of learning (summative) (Stiggins 2002, Chappuis and Chappuis 2002, and Davies 2002). Formative assessment (assessment for learning) should focus on the student since it provides effective and timely feedback to students, actively involves them in their own learning, and invites students' self-assessment of learning (Black and William, 1998b). Ideally, formative assessment should improve students' motivation and self esteem.

Some of the confusion for teachers has occurred, however, because there are significant differences between the assumptions of formative assessment and the stated purposes of the Grade Three Assessment. According to the Department of Education's policy manual, the primary purpose of the Grade Three Assessment is to provide "meaningful information to parents regarding their child's foundation

knowledge and skills in Reading and Numeracy" (METY, May 2002, p. 2, italics added). By providing information, parents and teachers can work collaboratively to assist students who need help. While the latter goal is consistent with those of formative assessment, the means for reaching it focus more on parents than students. As Dickens and Li have noted, *parents are best partners*. In fact it could be argued that the six principles of the assessment:

- assessment of student achievement is an integral part of teaching and learning
- the practice of assessment must incorporate a positive capacity building experience for teachers
- teachers need to be involved as the most important professionals in the assessment
- using a multitude of methods to assess students is a sound practice
- parents are best partners
- the search for alternative approaches to large-scale assessment (Dickens and Li, 2002, p.3-5)

focus more on parents and teachers than students. Not one of the principles speaks of improving student performance and motivation directly.

The result is confusion in the minds of teachers as to the purpose of the assessment. As mentioned earlier, Leanne clearly felt that the purpose of the assessment was to try and bring every student up to standard; as a result she assessed, re-taught, assessed again, and kept doing this until her students came up to standard. She was clearly focusing on the student. On the other hand, Margaret did little or no review with her students. She asked students to try to remember what they had done in Grade Two and to try their best on this assessment. She felt that she had to do this so that the results were "fair" (6, 120-124). One is left to wonder just how "fair" the results were if these two teachers used such different approaches to the assessment.

Another area that created confusion was the very term *assessment*. In most educational contexts, assessment means a "gathering information process about student achievement" (Chappuis and Chappuis, 2002, p. 149) that "informs teaching and helps students to learn more" (Davies, 2000, p. 1). The assessment certainly gathers information about student achievement, and most of the teachers felt that it informed teaching. But does the Assessment help students to learn more? When asked whether or not the Assessment was an effective way to provide feedback to students, seven

teachers said it was not. The information it gathered was mainly for parents and teachers.

In another way, in fact, the Manitoba Assessment is more like evaluation than assessment. Evaluation involves "reviewing the evidence and determining its value" (Davies, *ibid*, p. 1)—in other words, making decisions or judgments about worth. In the case of the Grade Three Assessment, teachers must make a judgment on how well each student has learned and place that student on one of three levels of proficiency for each competency. This "evaluation" is then signed and sent home as a report to parents. By judging the level of performance that a student has reached, the teacher is in fact evaluating the student rather than assessing them. This fact made the teachers feel professionally responsible for their decisions and created their need for concrete evidence to support their decisions. That is also why so many of them assessed each competency using paper and pencil tasks that could be used as evidence to support the teacher's decision.

In a similar way, the supposedly "formative" Grade Three Assessment actually has many summative attributes, in that it summarizes what students have learned to a particular point in time (Chappuis and Chappuis, 2002, p. 154). A summative report evaluates students and assigns

grades based on the results. In the Manitoba Grade Three Assessment, students are assessed on each competency and then, based on the results, a level of proficiency (grade) is assigned. The report that goes to parents summarizes what students have learned to that point.

As mentioned earlier, Levin and Wiens note that there is still work to do so that the Grade Three Assessment "becomes an integral part of teachers' ongoing instruction as opposed to an externally mandated requirement that is unrelated to day-to-day practices" (p. 7). To help in this regard, perhaps the policy manual could give a clearer definition of key terminology (assessment, evaluation, formative, summative) and refocus the purpose for the assessment. This could go a long way towards quelling the misgivings that teachers have and help to make the assessment a more integrated part of teachers' ongoing instruction.

For many of the teachers in the study, important questions remained surrounding the data collected. They questioned the validity of results when diverse tools were used, when the time spent on each assessment was varied, when the amount of review time was varied, when the number of tries a student had on each competency varied, and when the amount of teacher assistance varied. Leanne summarized

this concern when she said, "it's [the assessment] not really standardized and yet it's trying to be a standardized test" (10, 432-433). Given the differences in how the assessment is administered, Margaret saw "no need for the results to be published" (6, 482). Gavin too, wondered just what was being achieved by publishing provincial results given that the difference in administration across the province would probably be even greater than within in his school division (2, 129). Again, a good deal of the confusion these teachers expressed could be alleviated by a review and revision of the purpose and terminology of the assessment. The observations about standardization seem to indicate that professional judgment has replaced standardization of procedures, assessment tools, and timelines. While professional judgment is desirable in many ways, it must be based on a thorough understanding of assessment literature and practice. Providing and maintaining such understanding is the challenge.

One of the department's principles of the Grade Three Assessment is that "teachers are the most important professionals in the assessment process" (Dickens and Li, 2002, 3-5). As the most important person in that process, the teacher must take responsibility for ensuring that his/her skills and knowledge are current. From the teacher

interviews it would seem that the division may want to consider providing additional professional development for teachers. By providing further professional development opportunities for teachers, divisions would be moving more towards another department principle of the assessment, that the "practice of assessment must incorporate a positive capacity building experience for teachers" (ibid, 3-5). This can be done through professional development which focuses on a complete understanding of how to use assessment as a teaching tool (Dickens and Li, p.p. 3-5).

Teachers wondered why the department seemingly randomly selected the beginning of Grade Three as the most desirable level for assessment. Most of the teachers favored moving the Assessment to the end of Grade Two, or even to the beginning of Grade Two or the end of Grade One. There were strong preferences for gathering accurate student assessments as early as possible in the student's educational life.

One is left to wonder why, if beginning of year (formative) assessment is important, why is Grade Three the only level in the Kindergarten to Grade Twelve educational system where this takes place? Furthermore, what is the role of the department in following up the Grade Three Assessment results for students who need help? Should the

department allocate extra resources or carry out a further assessment at the end of Grade Three or beginning of Grade Four to find out if the students who were in need of help are now performing at grade level? Supports for students who need help and all additional follow-up assessments are currently left in the hands of the teachers and at the school and/or divisional level. In fact, now that the Grade Six Language Arts and Senior One Math Standards Tests have been cancelled for the 2004/05 school year, the next time that students in Manitoba are formally assessed by the Department of Education is at the end of Senior Four.

References

- Assessment Reform Group (1999). *Assessment for Learning: Beyond the Black Box*, Cambridge, England. University of Cambridge School of Education.
- Barlow, M., & Robertson, H.J. (1994). *Class Warfare: The Assault on Canada's Schools*. Toronto, Ontario: Key Porter Books.
- Black, P. & William, D. (1998a). Assessment and Classroom Learning. *Assessment in Education: Principals, Policy and Practice*, volume 5, no. 1. London, King's College.
- Black, P. & William, D. (1998b). *Inside the Black Box: Raising standards through classroom assessment*. London, King's College.
- Chappuis, J. & Chappuis, S. (2002). *Understanding School Assessment: A Parent and Community Guide to Helping Students Learn*. Portland, Oregon. Assessment Training Institute.
- Crooks, T. (1988). The Impact of Classroom Evaluation Practices on Students. *Review of Educational Research*, 58, no. 4, 438-481. New Zealand, University of Otago.
- Davies, A. (2000). *Making Classroom Assessment Work*. Courtney, British Columbia. Connections Publishing.

- Dickens, W, & Li, J. (2002). *Changing Approaches to Large-Scale Assessment*. A paper presented at the Provincial Testing in Canadian Schools: Policy, Practice and Research: A Symposium on Large-Scale Assessment. Victoria, British Columbia.
- Earl, L.M. (1995, May). *Accountability in Education: Where Do Teachers Fit?* Paper prepared for the Canadian Teachers' Federation National Conference "Public Education: Meeting the Challenges", Montreal, Quebec, May, 1995.
- Earl, L., & Cousins, J.B. (1995). *Classroom Assessment: Changing the Face; Facing the Change*. Toronto: the Ontario Public School Teachers' Federation.
- Froese-Germain, B. (1999). *Standardized Testing: Undermining Equity in Education*. Ottawa, Ontario: Canadian Teachers' Federation.
- Hansard (May 14, 2002). Oral Question Period in the Manitoba Legislature.
- Human Resources Development Canada. (2000). *Measuring Up: The performance of Canada's youth in reading, mathematics and science: OECD PISA study*. Council of Ministers of Education, Canada.

- Johnson Jr. J.F., Treisman, U., and Fuller, E. (2000).
Testing in Texas. *The School Administrator*, December 2000,
pp.20-26.
- Kohn, A. (2000). *The Case Against Standardized Testing:
Raising the Scores and Ruining the Schools*. Portsmouth,
New Hampshire, Heinemann.
- Levin, B. (1998, June). "An Epidemic of Education Policy:
(What) Can We Learn From Each Other?" *Comparative
Education*, 34(2), pp.131-141.
- Levin, B. and Wiens, J. (2003, May). School Reform That
Works: There is Another Way: A Different Approach to
Education Reform. *Phi Delta Kappan*, v84, i9, p.658.
- Leithwood, K. (2001). 5 Reasons Why Most Accountability
Policies Don't Work (and What You Can Do About It). *Orbit*,
Vol. 32, No. 1. Toronto, OISE.
- Manitoba Education and Training (1995, January). *Renewing
Education: New Directions, The Action Plan*. Winnipeg,
Manitoba. Clayton Manness, Minister.
- Manitoba Education and Training (1995, June). *Renewing
Education: New Directions, A Foundation for Excellence*.
Winnipeg, Manitoba. Linda G. McIntosh, Minister.

Manitoba Education, Training and Youth (2001, September).

Your Child and the Grade Three Assessment: A Brochure for Parents. Winnipeg, Manitoba. Drew Caldwell, Minister.

Manitoba Education, Training and Youth (2002, May). *Grade 3 Assessment in Reading, Lecture and Numeracy/Notions de calcul and Grade 4 Assessment in French Immersion Lecture.* Winnipeg, Manitoba. Drew Caldwell, Minister.

Manitoba Education, Training and Youth (2002, September). 2002-2003 Senior 4 English Language Arts Standards Test. *Information Bulletin.* Winnipeg, Manitoba. Drew Caldwell, Minister.

Manitoba Education and Youth (2003, June). Grade 6 English Language Arts, English LA-Immersion and Anglais Standards Test. *Administration Manual for Supervising Teachers.* Winnipeg, Manitoba. Ron Lemieux, Minister.

Manitoba Education and Youth, Assessment and Evaluation Branch (2003). *Grade 3 Assessment-Provincial Report 2002-2003.* Winnipeg, Manitoba. Ron Lemieux, Minister.

Popham, W.J. (2001). *The Truth About Testing: An Educator's Call to Action.* Alexandria, Virginia, ASCD.

Provincial Testing: Achievement Tests. Retrieved October 5, 2003 from Alberta Learning Web site:
www.edc.gov.ab.ca/k_12_testing

- Schmoker, M. (1999). *Results: The Key to Continuous School Improvement*. Alexandria, Virginia, ASCD.
- Sergiovanni, T.J. (2000). Standards and the Lifeworld of Leadership. *The School Administrator*, September 2000, pp.6-12.
- Shepard, L.A.. (2001). The Role of Classroom Assessments in Teaching and Learning. *Handbook of Research on Teaching*, 4th edition, Richardson, V. (Ed.). Washington, D.C. American Educational Research Association.
- Stiggins, R.J. (1999). *Assessment, Student Confidence, and School Success*. Portland, Oregon. Assessment Training Institute.
- Stiggins, R.J. (2002). *Assessment, FOR Learning. A Vision for the Future of Assessment in the United States* Portland, Oregon. Assessment Training Institute.
- Wiggins, Grant P. (1993). *Assessing Student Performance: Exploring the Purpose and Limits of Testing*. San Francisco, California, Jossey-Bass.
- Wiggins, Grant P. (1998). *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*. San Francisco, California, Jossey-Bass.
- Winnipeg Teachers' Association, (2004). *A Good Idea Gone Bad*. A publication of the Winnipeg Teachers' Association.

Footnotes

1. Numeracy skills are the ability to work with numbers in various ways and to solve math-related problems (Manitoba Education, Training and Youth, September 2001).
2. The three assessed levels of performance are: "needs on-going help"; "needs some help to meet expectations"; and "meets expectations".
3. Standards tests in Grade Six Language Arts and Senior One Mathematics are optional.

Appendix A:**Interview Guide**

This interview guide formed the basis for the teacher interviews. The questions provided structure for the interviews. Although the conversation with each teacher was based on the written questions, many of the interviews also flowed naturally and followed the interests, views, and perspectives of the teacher being interviewed.

Initial Background Questions (for context and rapport building)

- o How long you have been teaching in general?
- o How long you have been teaching grade three students?
- o How many years have you taught at this school?
- o For how many years have you administered the Grade Three Assessment?
- o Tell me about the school to help me get a picture of the students, staff and community that it serves?
- o Tell me about your class. Are they single grade, mixed grade, how many students, numbers of boys and girls, special needs students, and unique characteristics?
Please do not tell me individual student names.

**Further questions grouped according to research on
assessment that may impact teachers:**

1. Increased understanding and use of assessment

- o What have been the major changes for you administering the Grade Three Assessment rather than the Grade Three Standards Test?
- o Has your knowledge and use of assessment increased by administering this assessment? If so, in what ways?
- o Is the assessment an improvement on the former standards test?

2. Integrated use of formative assessment

- o Is the Grade Three Assessment an effective way to provide feedback to students?
- o Were you able to integrate the assessment into your daily activities or did you have to make adaptations to your teaching practices and student activities?
- o If you did make adaptations to your teaching practices and student activities, please tell me about them.
- o How much revision/re-teaching do you do prior to assessing each competency?
- o Do you assess each competency separately or are you able to integrate the assessment into normal, daily activities? Please give examples.

3. The balance between summative and formative assessment

- o What do you understand by the term *summative assessment*?
- o What do you understand by the term *formative assessment*?
- o How do you use formative and summative assessment in your evaluation procedures?
- o How have you learned about good assessment practices? Did you take courses in University, have you attended professional development sessions, and do you feel supported by your school and/or division?

4. Diverse pedagogical methodology

- o Did all students take the same assessment at the same time?
- o Did you differentiate assessment strategies to meet the needs of different children?
- o Did department or school division suggested strategies encourage you to try new ideas? If so, can you give me a few examples to show how you put these ideas into practice?

5. Changes in teacher-student relationships

- o How well do you feel you knew the students before you assessed them?
- o How did your children react to the literacy assessment?

- o How did your children react to the numeracy assessment?
- o How did your students react to you during the assessment?
- o Was this typical of their reactions to you on a day-to-day basis?

6. Changes in how and when curricular content is delivered

- o How would you like to begin your year with your new class of students?
- o Were you able to begin your year in this way or did the assessment cause you to make changes?
- o Did the assessment cause you to change the way you introduced topics/curricular content, or the length of time you spent on those topics/curricular content?
- o This assessment evaluates Grade Two outcomes. How did you balance teaching Grade Three outcomes at the same time as you were assessing Grade Two outcomes?

7. Increased teacher knowledge of student performance

- o In general, is the assessment an appropriate way to find out if students are progressing well?
- o What new information, if any, did you learn about your students?

- o If you did learn new information, how did you use the new information you learned to change your classroom practice?
- o Were additional supports put in place after the assessment for students who did not do well? What kind of supports were they? Were any these supports already in place before the assessment?

One final question:

Do you have other things you would like to talk about concerning the assessment?

Appendix B:

Student Report: Reading Grade 3 Entry

School: _____

Student Name: _____ Date: _____

Reading Competencies	Levels of Performance		
	Needs on-going help	Needs some help to meet expectations	Meets expectations
Reflection Student's ability to think about own learning as a reader	<input type="checkbox"/> Student begins to participate in guided reflection and goal setting to plan "What can I do" to be a better reader.	<input type="checkbox"/> Student participates actively in guided reflection on strategies and goal setting by talking about "what I did that helped me" be a better reader.	<input type="checkbox"/> Student reflects on strategies and goal setting by talking about "what I did that helped me" be a better reader; and sets realistic reading goals.
Oral Reading Skills and Strategies Student's ability to use a variety of strategies to read	<input type="checkbox"/> Student reads below grade level text and is unable to self-correct	<input type="checkbox"/> Student reads grade level text with some assistance to self-correct.	<input type="checkbox"/> Student reads grade level text and self-corrects when meaning is unclear.
Reading Comprehension Student's ability to understand and draw conclusions from text	<input type="checkbox"/> Student demonstrates limited understanding of what is read and makes some personal connections and assistance.	<input type="checkbox"/> Student demonstrates some understanding of what is read and makes some personal connections.	<input type="checkbox"/> Student demonstrates understanding of what is read and can explain personal connections.
Comments:			

In accordance with Manitoba Education, Training and Youth policy, the purpose of this assessment is to inform Parents/guardians of their child's level of performance on identified critical competencies in Reading.

Reading Assessment

Teacher:

Parent/Guardian:

Teacher's Signature_____
Parent's/Guardian's Signature

- I would like to discuss your child's performance. I will contact you to set up an appointment.

- I would like to discuss my child's performance.

Please sign and return to the school.

Appendix C:

**Report to Division:
Numeracy – English Program
Grade 3 Entry**

Report the number of students in each cell

Numeracy Competencies	Levels of Performance			
	Needs on-going help	Needs some help to meet expectations	Meets expectations	*Total # of students
Student sorts objects using one mathematical attribute; identifies attributes such as <i>shape</i> and <i>size</i>				
Student selects the appropriate standard unit (cm, m); estimates and measures length				
Student recalls addition facts to 10				
Student recalls subtraction facts to 10				
Student represents and compares numbers, using terms such as <i>even, odd, more, less, same as</i> , to 100				
Student understands place value to 100				
Student identifies, extends, and describes mathematical (repeating and growing) patterns				
Student solves and creates addition and subtraction story problems				
Student reads and interprets graphs				
<p>Please enter the number of students not included in reporting of competencies _____ Please explain.</p> <p>*Please explain any discrepancy in the "Total # of Students" column</p>				

Appendix D:

**Report to Division:
Reading – English Program
Grade 3 Entry**

Report the number of students in each cell

Reading Competencies	Levels of Performance			Total # of students
	Needs on-going help	Needs some help to meet expectations	Meets expectations	
Reflection Student's ability to think about own learning as a reader				
Oral Reading Skills and Strategies Student's ability to use a variety of strategies to read				
Reading Comprehension Student's ability to understand and draw conclusions from text				
<p>Please enter the number of students not included in reporting of competencies _____ Please explain.</p> <p>Please explain any discrepancies in the "Total # of Students" column.</p>				

**Please sign and return this form to the Admin Office
by November 15th.**

Number of students in Grade 3 (September 30 enrolment): _____

Number of reports sent to parents: _____

Appendix E:

**Report to Division: Numeracy – English Program
Grade 3 Entry**

Report the number of students in each cell

Numeracy Competencies	Levels of Performance			*Total # of students
	Needs on-going help	Needs some help to meet expectations	Meets expectations	
Student sorts objects using one mathematical attribute; identifies attributes such as <i>shape</i> and <i>size</i>				
Student selects the appropriate standard unit (cm, m); estimates and measures length				
Student recalls addition facts to 10				
Student recalls subtraction facts to 10				
Student represents and compares numbers, using terms such as <i>even, odd, more, less, same as</i> , to 100				
Student understands place value to 100				
Student identifies, extends, and describes mathematical (repeating and growing) patterns				
Student solves and creates addition and subtraction story problems				
Student reads and interprets graphs				
Please enter the number of students not included in reporting of competencies _____ Please explain. *Please explain any discrepancy in the "Total # of Students" column				

**Please sign and return this form to the Admin Office
by November 15th.**

Number of students in Grade 3 (September 30 enrolment): _____

Number of reports sent to parents: _____

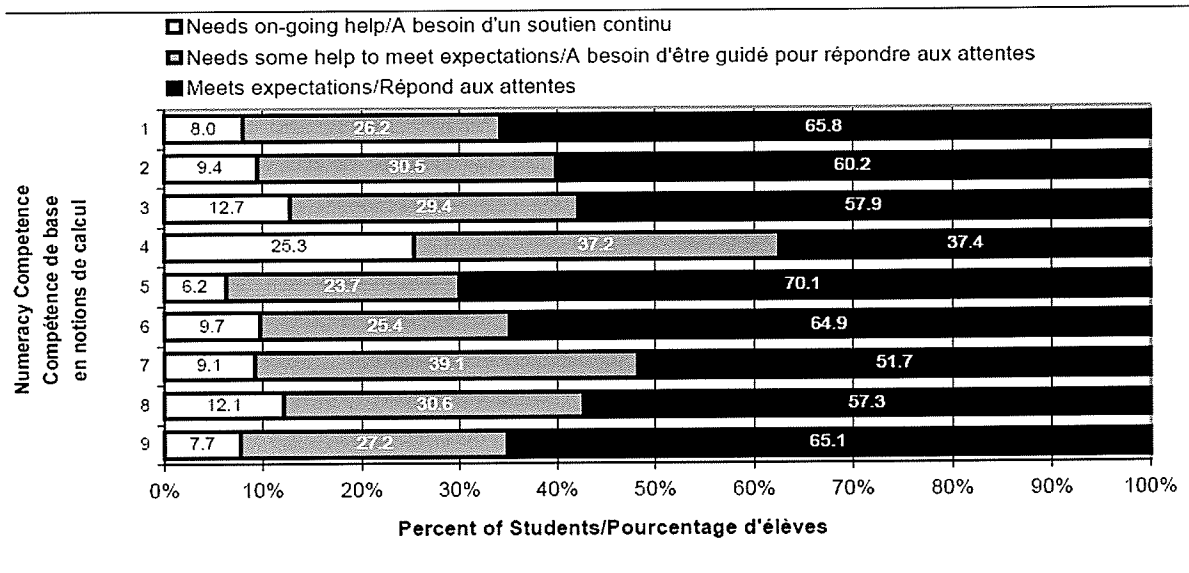
School

Date

Principal's signature

Appendix F:

Figure 1: Provincial Summary Results
Grade 3 Numeracy—English, French Immersion, and Français
Programs

**Numeracy Competency**

1. Student sorts objects using one mathematical attribute; identifies attributes such as *shape and size*
2. Student selects the appropriate standard unit (cm, m); estimates and measures length
3. Student recalls addition facts to 10
4. Student recalls subtraction facts to 10
5. Student represents and compares numbers, using terms such as *even, odd, more, less, same as*, to 100
6. Student understands place value to 100
7. Student identifies, extends, and describes mathematical (repeating and growing) patterns
8. Student solves and creates addition and subtraction story problems
9. Student reads and interprets graphs

Appendix F:

Figure 2: Provincial Summary Results by Program
Grade 3 Numeracy-English Program

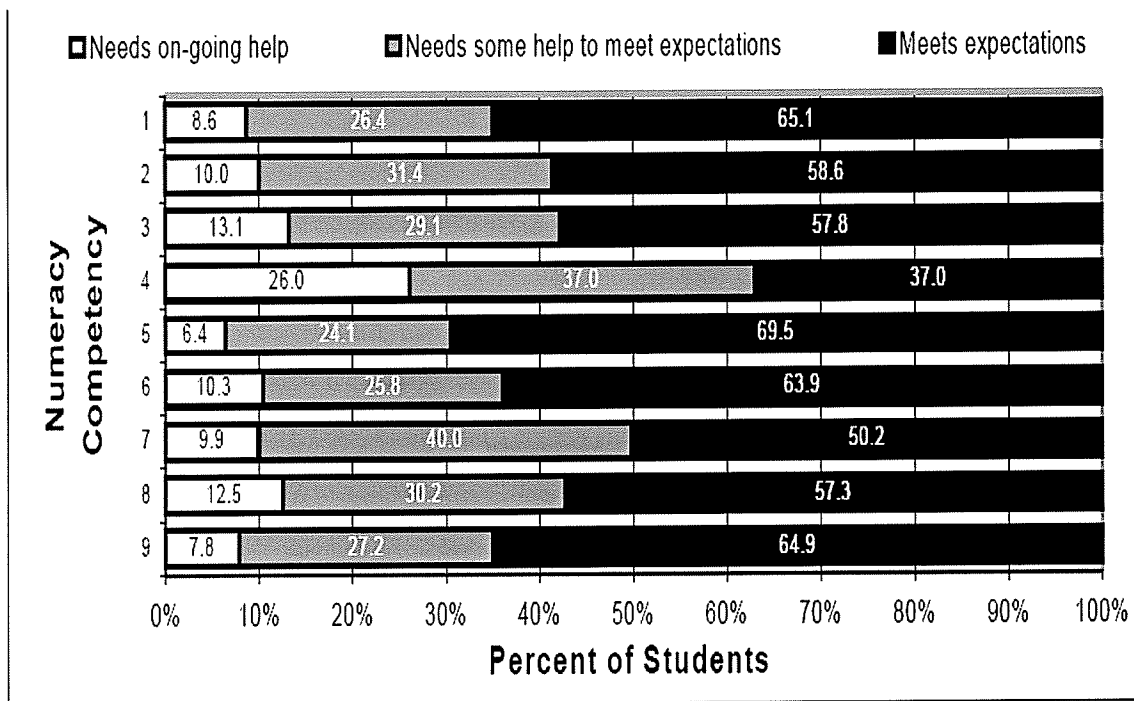
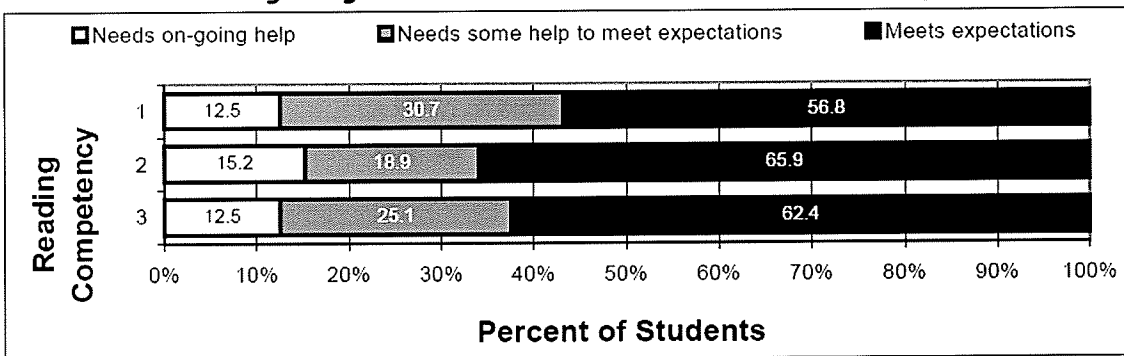


Figure 5: Provincial Summary Results by Program
Grade 3 Reading-English and French Immersion Programs



Reading Competency:

1. Reflection

Student's ability to think about own learning as a reader.

2. Oral Reading Skills and Strategies

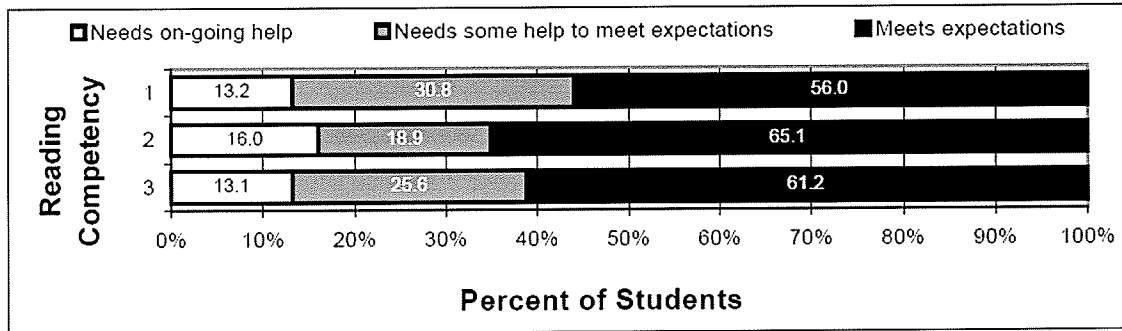
Student's ability to use a variety of strategies to read.

3. Reading Comprehension

Student's ability to understand and draw conclusions from text.

Appendix F:

**Figure 6: Provincial Summary Results by Program
Grade 3 Reading-English Program**

**Reading Competency:****1. Reflection**

Student's ability to think about own learning as a reader.

2. Oral Reading Skills and Strategies

Student's ability to use a variety of strategies to read.

3. Reading Comprehension

Student's ability to understand and draw conclusions from text.