THE ASSESSMENT-CURRICULUM RELATIONSHIP: CONSEQUENCES FOR TEACHER INSTRUCTION AND STUDENT ASSESSMENT

by

Ronald Ernest Saranchuk

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Department of Curriculum, Teaching and Learning
Ontario Institute for Studies in Education of the University of Toronto

© Copyright by Ronald Ernest Saranchuk, 1998
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-35310-9
The Assessment-Curriculum Relationship:  
Consequences for Teacher Instruction and Student Assessment  
Ron Saranchuk, Ph.D., 1998  
Department of Curriculum, Teaching and Learning  
University of Toronto  

Abstract  

This interpretive case study compared the relationships between assessment and curriculum, and their consequences on teacher instruction and student assessment in the classrooms of two research sites. One research site was in an early stage, the other was in a more advanced stage of use with an assessment innovation (the Toronto Benchmark Program). The Program emphasized mastery learning, and standards of performance at key stages in students’ education.  

In the early site, where the curriculum was found to be content-driven, the assessment procedures had limited interaction with classroom instruction. Assessment practices were mainly summative, and instruction was growth-centred. In the advanced site, where the curriculum was standards-driven, the assessment and instructional procedures interacted with each other to a greater extent. The assessment was mainly formative, and instruction was information-centred. Simultaneously, data emerged that clarified how contextual, and background factors influenced classroom use of the innovation.  

This study may be relevant for practitioners as well as researchers since technical and management dimensions of change have been portrayed through the “first-voice” experiences of those most closely associated with the development, implementation, and consequences of an assessment innovation.
Acknowledgements

The advice and encouragement of Dr. Joel Weiss, my thesis advisor, have been invaluable to me in this research study. His many hours of support in helping me refine my research question, develop my research design, and relate my research findings to the more global issues associated with assessment and curriculum were much appreciated!

My two other committee members, Dr. Phil Nagy and Dr. John Ross, also gave of their time and expertise in suggesting possibilities for focussing on assessment and curriculum issues pertinent to teachers as well as to the academic research community. Along with these members, Dr. Brad Cousins, Dr. Lorna Earl, and Dr. Merle Wahlstrom initiated insightful and constructive dialogue during my oral defence on matters relating to assessment and curriculum.

I am particularly indebted to the participants in this study. Not only did they welcome me into their world of work for an extended period of time, but they also shared with me their experiences and reflections on how they developed and implemented their assessment-curriculum innovation in their classrooms.

I am also indebted to my wife, Yvette, who offered her insights into the issues associated with the complex interplay between assessment and curriculum.
# Table of Contents

Abstract                                      ...ii  
Acknowledgements                             ...iii  
Table of Contents                             ...iv  
List of Tables                                ...vii  
Appendices                                    ...viii

Chapter I                                     
The Study: Assessment and Curriculum in Change ...1  
1. Introduction                               ...1  
2. Issues Pertinent to this Study             ...3

Chapter II                                    
Prologue                                     ...10

Chapter III                                   
The Problem Stated                           ...18  
1. Research Objectives                       ...20  
2. Theoretical Assumptions                    ...21  
3. Delimitations and Limitations              ...22  
4. Significance of the Study                  ...23  
5. Terminology                                ...24

Chapter IV                                    
Review of the Literature                     ...27  
1. Educational Change                        ...27  
2. Classroom Assessment                      ...38  
   i. Traditional                             ...38  
   ii. Performance                            ...40  
3. Consequences of Classroom Assessment      ...42  
   i. Traditional                             ...42  
   ii. Performance                            ...46  
4. Research Relevant to this Study           ...50  
5. Summary of Research                       ...58
Chapter V

Methodology

1. Research Design ...59
2. Research Rationale ...61
3. Access ...64
4. Ethical Considerations ...65
5. Participants ...66
6. Data Collection Methods and Instruments ...68
   i. Interview ...70
   ii. Observation ...72
   iii. Documentary Analysis ...73
   iv. Researcher Field Notes ...74
   v. Member Checks ...75
7. Interpretation ...75
8. Data Analysis ...79

Chapter VI

Context of the Montrose and Arcadia Research Sites ...82
1. Montrose and Arcadia Research Sites ...82
   i. Introduction to the Innovation ...85
   ii. Administrative Experience ...86
   iii. Professional Concerns ...88
   iv. Documents Used ...91
   v. Approach to Student Performance ...93
2. Shift to Performance Assessment ...95
3. The Innovation ...101

Chapter VII

Findings of the Study ...106
1. Antecedents - Setting the Stage for Mobilization ...106
   i. Conceptualization ...106
   ii. Adoption ...111
   iii. Process ...115
   iv. Support ...118
   v. Communication ...121
   vi. Staff Turnover ...123
   vii. Teacher Resistance ...123
List of Tables

1. Factors Associated with Development ........31
2. Determinants of Development from the Eight Year Study ..........32
3. Variables Influencing Implementation ........33
4. Factors Affecting Implementation and Continuation ........35
5. First Framework for Interpretive Research ..........76
6. Second Framework for Interpretive Research ..........77
7. Third Framework for Interpretive Research ..........78
8. Antecedents to Curriculum Change ..........175
10. Assessment-Curriculum Outcomes ..........185
## Appendices

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.</td>
<td>Teacher Interview Schedule</td>
<td>231</td>
</tr>
<tr>
<td>B.</td>
<td>Student Interview Schedule</td>
<td>232</td>
</tr>
<tr>
<td>C.</td>
<td>Administrator Interview Schedule</td>
<td>233</td>
</tr>
<tr>
<td>D.</td>
<td>Parent Interview Schedule</td>
<td>234</td>
</tr>
<tr>
<td>E.</td>
<td>Classroom Observation Schedule</td>
<td>235</td>
</tr>
<tr>
<td>F.</td>
<td>Toronto Board of Education Informed Consent</td>
<td>236</td>
</tr>
<tr>
<td>G.</td>
<td>Administrator Informed Consent</td>
<td>237</td>
</tr>
<tr>
<td>H.</td>
<td>Teacher Informed Consent</td>
<td>238</td>
</tr>
<tr>
<td>I.</td>
<td>Parental Informed Consent</td>
<td>239</td>
</tr>
<tr>
<td>J.</td>
<td>Toronto Benchmark Performance Criteria</td>
<td>240</td>
</tr>
</tbody>
</table>
Chapter I

The Study: Assessment and Curriculum in Change

1. Introduction. This study is grounded in the assumption that a clear understanding of the meanings, roles and consequences on school curricula and practice inherent in the assessment-curriculum relationship is foundational to teachers' instruction and assessment practices. Critical to the design of this study, the Toronto Benchmark Program is viewed in two schools that are in varying stages of development and use with the Program. This program is designed to be, simultaneously and seamlessly, an assessment as well as a curriculum program. Consequently, the understanding of the impact of assessment-curriculum relationship(s) on classroom practice, in the two research sites in this study, is necessarily intertwined with an understanding of issues dealing with curriculum change.

The relationship between assessment and curriculum may be either assessment-driven or curriculum-driven. Popham (1983) explains that assessment-driven curriculum (ADC) occurs when a high-stakes test (Mehrens & Popham, 1992) of educational achievement influences the instructional program that prepares students for the test. For example, because teachers want or need students to do well on tests that have important contingencies associated with them (such as certification and placement based on final examinations) they may focus a significant portion of their instructional activities on the knowledge and skills assessed by such tests (Madaus, 1991). Alternatively, researcher (Nitko (1995a; Tyler, 1949) holds the view that in curriculum-driven assessment (CDA) the curriculum, not the assessment tasks,
should be the master of the assessment enterprise; that each step in the assessment development process must improve assessment validity; and that the development process must produce assessments that are seamless in both the teaching and the curriculum. It is the curriculum that provides the steps helping to move students' development from the subject content to the assessment procedures.

Within the framework of educational change, the technical aspects of the Benchmark Program (as an assessment innovation) are viewed closely in this study. As noted by Broadfoot (1994), educators, parents and policymakers are continuing to ask for educational assessments that serve two seemingly incompatible purposes: 1). to determine whether students are achieving or striving toward desired standards of performance; and 2). to provide relative measurements of student achievements within and across classrooms, schools, school boards, provinces, and countries. These requests call for a complementary relationship between assessment and curriculum. Examples of efforts that reflect mergers between assessment and curriculum are reforms such as the California Assessment Program (California Assessment Committee, 1991), the developments recommended by the Task Group on Assessment and Testing in England and Wales (TGAT, 1988), the Dutch Foundation for Educational Research (SVO) in cooperation with the Central Institute for Test Development (CITO), the Policy on Curriculum and Assessment in Scotland (Committee on Assessment, 1990), the Toronto Board of Education's Benchmarks: Standards of Student Achievement (Rutledge, 1993), and the Australian (Victoria) Curriculum and Standards Framework (Board of Studies, 1994). All of these reforms have attempted to develop more "in context" (Gardiner, 1992) and "authentic" approaches to assessment
(Hill, 1993). Pertinent to this study, these reforms reflect the emphasis placed on the interdependence between assessment and instruction (Wood, 1986).

Although there are many descriptions and suggestions found in the literature about why teachers might change their assessment practices, the search revealed that very few studies have evaluated the impact of using teachers' [externally] mandated classroom assessment practices on their instruction. As a result, we have little knowledge of the consequences on classroom practices that are associated with a program's background, development, and use. This research describes the complex, creative and dynamic adaptations teachers use to introduce change to their classrooms. It also views teachers', students', administrators' and parents' perspectives and explanations of their classroom practices as central to their understanding of the change process. Such information is crucial to, but missing from, calls for educational change (Spencer, 1996).

The consequences of classroom assessments are first articulated through the voices of the participants involved in the study, and then interpreted within the broader framework of assessment and curriculum issues found in the literature. These interpretations shed light on how the relationship between assessment and curriculum affects teachers' instructional and assessment procedures.

2. Issues Pertinent to this Study. Issues related to curriculum change concern the change process itself, conceptualization of the change, control over the curriculum, teacher role in the educational change, and the development of curriculum.

Educational change can be interpreted in different ways by different people: it may be viewed as fidelity (Fullan, 1975), mutual adaptation
(Berman & McLaughlin, 1979), enactment (Eisner, 1985), and school-based (Brady, 1990). Underlying these approaches are assumptions that frame the change process on a continuum from replication to adaptation to local construction to local management; from curriculum as externally to internally controlled; and from those involved as subjects to receive the curriculum to those as participants to shape and process it. The approaches make assumptions about change that serve to separate them from each other. For instance, those assuming that change takes on a rational, systematic, linear process would be most comfortable with a fidelity perspective of change. Those viewing change as a less linear process might readily adopt the mutual adaptation perspective. Others who see change as entirely a process for growth for teachers and students - a change in thinking and practice, rather than an organized procedure for the design and the implementation of an innovation, might adopt the enactment perspective of implementation. Still others believe that the foundation of change is the organizational culture that supports it. Accordingly, one's assumptions of the change process may color one's meaning of change, and one's meaning of change may influence whether or not one adopts the change, and the way one might put it into use.

In particular, one's classroom instructional practices may also be closely linked to the change approach adopted. These practices may range from traditional, in which the teacher sets the pace, and determines the means and methods of instruction to the performance approach in which the child sets the pace, and is a major source for deciding on the means and methods of instruction. Teachers' instructional practices can differ (Raphael, 1993) according to several issues associated with each approach. Teacher role in the classroom may vary on a continuum from conveying knowledge to enabling students to process knowledge; classroom communication may vary in [uni or
bi] directionality; and student performance may vary from additive to transformative. In either case, conceptualization of the innovation is often directly related to its use. How well these approaches are internalized by the classroom teacher can be revealed through classroom instruction. For example, it might be instructive for those adopting the Toronto Benchmark Program to conceptualize that the program basically functions by merging behavioral objectives (Tyler, 1949) with levels of learning (Bloom, 1956). Such conceptualization grounds understanding for teachers and students of the day-to-day classroom operation of the Benchmarks since it informs how the program's classroom instruction and assessment work together. In large part, this issue speaks to a professional development concern. Teachers, and administrators would likely need time and theoretical support in conceptualizing the focus of the new program; changes to its purposes and practices; and new materials to be developed. This often requires well-planned, sustained and resourceful support in helping teachers exercise their central role in implementing new assessment programs (Stiggins, 1991). Without an appropriate affinity to the assessment and curriculum approach, change may be more nominal than actual (Romberg & Price, 1982).

Another issue often associated with the quality of change relates to control over the curriculum, i.e., whether the source of curriculum control resides in external provincial ministries, in state departments, in school boards, or in local teacher interests. Installing one curriculum to replace another may reflect particular political values that dominate the current educational context. While the political values mandate what teachers will do, the educational concerns of what ought to be done may receive low priority. In turn, indicators of educational success come to be seen as those which are quickly developed and readily measurable (e.g., declarative knowledge) while
problem-solving (e.g., procedural cognitive development) may be de-emphasized. A potential problem with the development and implementation of curricula by sources external to the teachers and students may be that the unique contexts of each classroom might not be taken into account by the developers. For example, external influences may emphasize behavioral outcomes which, in turn, may presume that students have the prerequisite skills to engage in tasks leading to attainment of the desired outcomes.

The consonance between teachers' values and the assumptions of traditional and performance-based assessments help to influence teachers' instructional practices. For example, if a teacher placed a high value on how a student's performance compared to the mean of those in his/her class, assessment for measurement might be preferred. Alternatively, if a teacher placed a high value on the gains made by an individual student in terms of a stable set of criteria over time, performance assessment might well be preferred over traditional assessment methodologies.

Researchers (Fullan, 1982; Neale, Bailey and Ross, 1981) conclude that curriculum change models depict stages that fall into four broad categories: initiation, implementation, integration, and outcomes. The former category (i.e., initiation) is especially critical to this study in that it provides added background helping to illuminate how the Toronto Benchmark Program was put into practice in two research sites. As MacDonald and Walker (1976) comment:

The question of how new ideas and practises spread from their point of origin and gain widespread adoption is central to any system of planned change. (pp. 4-5)

For instance, teachers who spend time and activities in preparing to operationalize an innovation at the classroom level, would more than likely be
taking into account the characteristics of the local environment within which they and the program must operate - a process that Berman and McLaughlin (1976) call mutual adaptation. Implementations in which only very little or only unilateral development occurs prior to putting an innovation into use may be forced either to fit into or to re-shape the context and run the risk of failure. Clearly, the combined development and implementation may illuminate the outcomes from the classroom use of the innovation.

Issues associated with classroom assessment in this study are the competing goals of assessment, the means of comparing student performance, the emphasis on assessment and/or curriculum, and the communication of assessment results.

Assessment practices can be closely associated with two legitimate, but different goals - the goal of measurement, and the goal of instruction. While teachers want testing and assessment to be designed to inform instructional decision making, administrators and school trustees often want testing and assessment to have optimal measurement properties. The issue is aggravated further when agencies external to the classroom (e.g., provincial governments) may control and mandate particular approaches to assessment. Standardized testing is usually mandated by those external to the classroom, and performance testing by those directly involved with the classroom setting. Because of the competitive nature of the views, there has often been very little integration between them.

Closely related to the issue of competing goals is that of comparing student performance. At times, there is a tendency to assume that comparison of student performance must necessarily be made against the middle-ranked score in a wider population of students. This form of comparison, i.e. through
norm-reference testing, differentiates between students on the basis of the median score on a normed test, and is used principally for purposes of selection. Far different is comparing student performance against a set of external criteria, i.e., criterion-reference testing. The external criteria may be the dimensions of a particular curriculum studied with the score differentiating between what was maximally expected, and what was actually achieved by each student. Such tests set criteria according to requisite performance levels, and is used to diagnose and guide student progress. However, criterion and norm referencing are closely tied to two differing views of educational purpose: the first sees education as a means for sorting students by common standards for future educational opportunities; the second is seen as an efficient means to track and select students (Darling-Hammond & Ancess, 1994).

One of the most critical issues in this study concerns the resulting "mix" produced from merging assessment for measurement and assessment for instruction. The mix produced involves merging of assumptions associated with assessment for measurement, and assessment for instruction (Taylor, 1994). On the one hand, assessment for measurement assumes that humans differ from one another on various human traits; that one person's measurement on a trait can be reported relative to the distribution of other similar individuals' measurements on that trait; and that instruments can measure these individual differences reliably. On the other hand, assessment for instruction suggests that we can set educational standards and strive toward them; that most students can internalize and achieve the standards; that very different student performances still reflect the same standards; and that educators can be trained to internalize the standards, and judge diverse student performances consistently and fairly. Depending on which set of assumptions
or which combination of assumptions one adopts, immense impact may be exerted on the meanings, and consequences of educational practises within classrooms.

In turn, the issue of communicating assessment and curriculum results become tied to how they might be of use for measurement and/or instructional purposes. In order to be of practical value to those for whom the results are intended, the findings need to be reported clearly, and accurately. The importance of fulfilling this goal is that students' strengths can be developed and their problem areas may be improved; that reporting may provide crucial information for conferences between teachers and parents/guardians, and students; and that transfer of assessment information within or between schools may be complete and resourceful to those receiving it.

For purposes of organization and clarity, the issues that fall under "curriculum change" and "classroom assessment" have been treated separately in this section. In the actual experience of implementing an assessment and curriculum innovation within the context of a classroom, the issues linked to classroom assessment are inextricably intertwined with those of curriculum change. Taken together, these issues help to provide the broader parameters within which the findings that evolved from the research question in this study can be interpreted. The perceptual lens through which the researcher interpreted the data collected in this study has been shaped by his personal and professional background experiences identified in the prologue of this report.
Chapter II

Prologue

The claims made regarding the two major issues in this study, i.e., on curriculum change and classroom assessment, are ultimately grounded in my own personal and professional beliefs. Therefore, in order to provide the reader with some means of perspective on the text of this study, it is necessary that I identify and clarify my background and beliefs.

During my eighteen years of classroom teaching, I spent considerable time developing a wide range of curricula for students in our public schools. Almost always, the students in my classrooms were of diverse interests and abilities, grade levels, and socio-economic groups. Because my main focus was on improving their classroom performance, I was most preoccupied with what they needed to learn, what they actually learned, how they learned it, and how I could feedforward this information to the instruction they were to receive from me. In essence, I was concerned with how I could integrate my assessment and curriculum in order to help my students improve their classroom performance.

During my professional teacher training, I was only vaguely aware of the specific relationship that existed between assessment and curriculum: the curriculum was to be taught; the students were to be tested; and the results were to be communicated to my principal, and to the parents of my students. At this time, my classroom practice and perception of this relationship was unidirectional: from curriculum to assessment! I gradually began to feel that I could do a lot more to enrich my classroom instruction.
I began to converse with other teachers and also began observing how they taught and assessed their students. My views and practices associated with assessment and curriculum began to change. I started seeing more scenarios that addressed what and how students were taught, often within the same grade and subject areas. Some teachers emphasized writing, reading, listening, and speaking activities in varying degrees while others attempted to balance the time devoted to these activities. Others lectured and utilized a combination of group and individual seat work, and developed student work stations for particular sets of classroom activities. What struck me most was that I saw only a limited number of ways in which teachers assessed their students' classroom performances. More particularly (except for end-of-term reporting to parents) I saw little use made of the information generated from my colleagues' classroom assessments. While some of them assessed their students' performances on a continual basis, others did so only at the end of major units of classroom activities. Almost always, the classroom assessments served measurement rather than instructional purposes. At the same time, most of these assessments were of the pencil-and-paper variety. Usually they measured students' identification and recall rather than application and judgment levels of learning. The entire educational process seemed to consist of delivering the mandated curriculum, and then of assessing students in order to find out how well they "received" the material delivered.

However, in a few classrooms, teachers' views and practices with assessment and curriculum appeared to be exceptional. To me, the essence of these "exceptions" was based not only on the many types of assessment activities utilized with students and on the increased amount of information generated from the assessments, but also on the fact that this information was also used to inform further instruction and performance! To me, the social
and educational activities in these classrooms combined to create an atmosphere of excitement. The students were almost always engaged in activities that seemed challenging and interesting to them. I was very intrigued by the way in which the teachers of these exceptional classrooms interacted freely and resourcefully in helping their students perform their classroom tasks. While classroom discussions focussed on assignment expectations, students dialogued enthusiastically about their lessons with each other, and with their teachers. The teachers made it a priority to be aware of the level in which each of their students was working. On a regular basis, students were given subject criteria and activities that helped them move to a higher level of performance. The students always seemed to know where they stood with regard to their classroom performance. Alternatively, teachers were aware of the subject criteria that served as the basis for their classroom instruction. I thought the ethos of these classrooms to be one of engagement and growth. Further, these exceptional classrooms demonstrated to me how assessment and curriculum activities could be integrated in actual practice, to enhance student performance.

I sensed that the regular and exceptional classrooms differed along many lines. The regular classrooms were teacher-centred, information-based, and assessment-driven. These classrooms were usually very quiet, with very little social interaction, much individual seat work, and with teacher-directed text and/or work-sheet assignments. There was little evidence of students working well with each other. In contrast, the exceptional classrooms were far more student-centred, performance-based, and curriculum-driven. There was much animated and purposeful conversation amongst students and the teachers. Students worked at several activity centres in the classrooms, often had a hand in choosing their own learning material, and did a lot of writing each day. It
was obvious that the students worked very well with each other. Not only were there more curriculum resources in the exceptional classrooms, e.g., supplementary student learning materials, but students were also engaged in more varieties of classroom activities. As well, there were more types of assessments in these classrooms with the result that teachers had access to a wider range of student performances. In turn, teachers in the exceptional classrooms had more feedback on which to base decisions for extended student growth. The assessments, curriculum, and student performances were integrated and supportive of each other whereas in the regular classrooms, these components stood by themselves to a greater degree.

I began, in more depth, to reflect on the meaning of curriculum and assessment. I recognized what is commonly referred to as the intended or mandated curriculum, i.e., the syllabus usually found in provincial or school statements and materials. Although I saw a lot of this type of curriculum in the regular classrooms, I saw more of a constructed or developed curriculum in the exceptional classrooms. The main difference between the two types was that the 'constructed' curricula addressed the abilities and interests of students particular to each classroom while the mandated curriculum provided a prescribed set of knowledge, skills, and values (regardless of students' abilities and interests). The former was designed to integrate while the latter was designed to separate context from classroom instruction. The operational curricula were almost always developed by the classroom teacher. In other cases, there were parts of Language Arts curricula that teachers would simply not teach and assess for personal or professional reasons. I became aware of the 'hidden' component in curriculum: in the absence of clear intentions for learning, and criteria for performance, students focussed on material they
thought might be on the classroom tests, and ignored material they thought might not be covered by the test.

My view of curriculum continued to evolve through studies of researcher-theorists such as Bloom, 1956; Dewey, 1916; Krug, 1956; Schwab, 1970; and Tyler, 1949. I see curriculum to be a means and a rationale through which schools coordinate educational experiences, materials and instruction. The means and rationales, in turn, guide schools in providing resources, and in creating environments conducive to student performance. I feel this description of curriculum is sufficiently comprehensive to allow for classroom experiences to be planned or incidental, and to be operational or hidden. I now see curricula as the blueprint for providing students with learning opportunities, and this blueprint can be communicated explicitly to those involved with it. The most appealing aspect of this view is the [educationally] workable emphasis on "experience" as the basis for students' classroom performances. I feel this allows teachers and students opportunities to construct an environment in which students are engaged in experiences that enhance the development of their classroom performance. My constructivist bias towards curriculum is a chord that runs through the text of this inquiry from beginning to end.

At the same time, I reflected on practices in which assessment was used mainly for summative rather than for diagnostic and formative feedback; for emphasizing behavior and content rather than thought and concepts; for pupils as passive recipients rather than as active participants of assessment; and for the practice of aggregating rather than attending to pupils' separate classroom achievements. Eventually, I came to view assessment as data collection, i.e., as the task of using a variety of techniques in gathering information about one or more students. To me, assessment describes student performance while
evaluation judges it after integrating the entire collection of students' performance data from diverse sources. Thus, in order to promote growth in students' classroom performances, I now see assessment and evaluation as necessary classroom practice, and as complementary to each other in the educative process. As well, I realized that views of assessment could differ according to the rewards and sanctions associated with it. For instance, when the assessment data is used to make important decisions that directly and immediately affect students, as in entrance into university, it is viewed as "high stakes". Alternatively, when student performance on assessments is of no immediate and direct consequence, e.g., where results from a standardized test are not used to evaluate students, the assessments are viewed as "low stakes". What intrigued me was how the varying degrees of importance attached to assessments influenced the curriculum being taught. In addition, high and low-stakes views associated with student performances seemed to result in different kinds of consequences for classroom practice.

Although this reflection helped me think about the possibilities for merging assessment and curriculum. two thoughts remained with me. First, the relationship between assessment and curriculum could be reciprocal, i.e., each could influence the other. Second, I wondered what specific consequences might evolve from each relationship. I began examining my own assessment and instructional practices: what emphases did I place on assessment and on curriculum? During the first four years of my teaching, I assessed students to see if they knew the content of the course(s) I taught them, and to use the assessment results as the basis for deciding whether or not to promote my students to the next grade level. Clearly, I practised high-stakes assessment. As I developed my teaching abilities, I shifted from assessing my students for
purposes of measurement and selection; I assessed students more and more for purposes of instruction and program development. I found that I was collecting information that I could use not only to evaluate my students, but also to plan for instruction that was focused on developing each student's performance. I became more involved in developing my Language Arts curriculum in order that it could provide more opportunities for student growth. Though I still used the results from my students' assessments to make decisions regarding their promotion to the next grade level, my emphasis had changed from a purpose of measurement for knowledge content and grade promotion to one of feedback that guided my students' performances, and my classroom instruction. Though I couldn't articulate precisely what had happened at the time, I had turned my Language Arts instruction around from being assessment-driven to curriculum-driven. Correspondingly, my use of assessments had shifted from high to low stakes. Although my work load increased with this shift, so did my students' enjoyment and performance in Language Arts.

As I was completing my doctoral coursework, I was motivated to pursue the study of assessment-curriculum relationships in a more systematic manner. I began reading about, and speaking to people who had direct experience in using the Toronto Benchmark Program. It dawned on me that this innovative assessment and curriculum program represented a vehicle for addressing my main critical concern: Could assessment practices and curriculum organization be combined in such a way to provide students in regular classes the educational benefits I experienced with some of my exceptional classes? Later, I came to realize that any educational program first needed to be developed, and put into use in teachers' classrooms before it might be of any value to the teachers and students working with it. Increasingly, I thought that
the process of conceptualizing, developing, adopting, and implementing a new program might be critically related to the success of a program.

In preparing my thesis proposal, I found extremely little research related to the study of particular cases involved with the interaction between assessment and curriculum within the context of educational change. For example, there are numerous studies on the effects of assessment on educational practices; and even more studies on educators' experiences with attempting to put educational programs into classroom practice. However, there is a dearth of studies that relate the specifics of classroom and school context to the change being examined. More specifically, no such studies have examined the consequences of an assessment program in which a period of development has preceded its implementation compared to another that has begun the change process without any preliminary preparation for change. Very few students, teachers, administrators and parents involved in this kind of interaction have ever had their "lived" experiences of the change documented meaningfully. This gap in the literature helped motivate me to conceptualize my research inquiry to include educational change in conjunction with examining the interaction of assessment and curriculum, and its consequences on classroom assessment, and teacher instruction.
Chapter III

The Problem Stated

In the past, the primary purpose of assessment in many education systems has been to select; to "weed out" the unfortunate majority who were needed to undertake low-level, unskilled jobs. This emphasis on assessment served well the mass education systems which accompanied the 19th century period of industrialization. However, today there is an increasing emphasis on describing learning outcomes in terms of standards achieved - often associated with the pre-specification of such outcomes in a way that reflects the integration of curriculum and assessment planning. However, any changes to the form or content of what is to be measured may bring about equivalent changes in curriculum emphasis.

Amidst a climate of growing concern for educational quality, various educational assessment program innovations (e.g., curriculum-based assessment, outcomes-based education, Ontario Ministry of Education provincial testing in Language Arts, Ontario Provincial Benchmark Program) have or are being implemented. Yet, despite the centrality of assessment in teaching, we know relatively little about how teachers integrate assessment with curriculum in our schools (Linn, 1990; Stiggins, Conklin & Bridgeford, 1986), how these relationships have evolved, and what consequences they have for teachers' instructional and assessment procedures. Further, much of the evidence on the precise effects of assessment on school curricula and instruction, though often based on the observations of people in close contact with schools, is notably scarce and informal (Madaus & Kellaghan, 1992).
Depending on the direction and intensity of the "drive" in the assessment-curriculum relationship, a myriad of consequences on school curricula and practices can emerge. While we are aware of some of the consequences on school curriculum and practice at the extremes of the CDA-ADC relationship (e.g., Nitko, 1989; Bowler, 1983), we know very little of the lived consequences when the nature of the assessment-curriculum relationship is not so clearly polemic. In addition, we have very little knowledge of the consequences of this relationship in interaction with the characteristics of the local educational sites in which they occur. As a result, severe limitations are placed on inferences made about a curriculum's actual potential for delivering on its goals within the contexts of its local site. Those concerned remain uninformed as to i) the curriculum and/or assessment emphases that drive instruction; ii) the consequences of this relationship on school curriculum and practice; and iii) the taking of appropriate action in ameliorating negative and/or promoting positive curricular outcomes.

The research in this study is designed to clarify, from the voice of the participants, the consequences that particular emphases on assessment and/or curriculum have on their classroom practices. In addition, the consequences are viewed in interaction with the characteristics of their local educational environments. One of the benefits of this study is that it may enable educators to make more reliable inferences, not only about a curriculum's capacity for fulfilling its goals, but also about the specific dimensions of its goals.

Accordingly, the overall problem in this study is stated as:
From the perspective of the participants involved in the Toronto Benchmark innovation, what are the consequences of assessment-curriculum relationships on teachers' instructional, and assessment procedures?

1. Research Objectives. Consistent with this case study's interpretive approach to research (Erickson, 1986), five objectives were used to guide this inquiry. These objectives relate to 1.) purposes and meanings; 2.) contexts; 3.) activities; 4.) relationships; and 5.) consequences.

1. To interpret the purposes and meanings attached to Benchmark assessment according to the students and teachers associated with it:

2. To identify the issues involved in the development of Benchmark assessment for curricula within schools of the Toronto Board;

3. To describe the nature of student-teacher classroom (and related) activities associated with Benchmark assessment:

4. To analyse how school curricula influence Benchmark assessment; and in turn, to analyse how Benchmark assessment influences its curricula; and

5. To examine how the consequences of Benchmark innovation affect teacher instruction, and assessment procedures.

While the preceding interview schedule serves to guide much of the data collection in this study, it did not determine the organization and emphasis placed on the data collected. In fact, the data collected were 1.) organized into
themes that, in some cases, cut across the specific divisions found in Erickson's interpretive framework; and 2.) treated with varying degrees of emphasis in this inquiry.

2. Theoretical Assumptions. An assumption basic to this interpretive study is that formal and informal social systems operate simultaneously, and that they are intertwined. In this study, the dimensions of role and status of teachers, students, administrators and parents help to form the essence of the social and cognitive organization of the classroom as an environment characterized with a particular relationship between assessment and curriculum.

Second, the focus on the processes and structures in social ecology (Lincoln & Guba, 1981) is intrinsic to understanding the ways in which people, in their actions together, constitute environments for one another. The classroom research sites in this enquiry are the foci for the construction of meanings-in-action which constitute the learning environment, the content to be learned, and the context in which student learning takes place.

Third, those living in their own environments have the capacity to construct their own norms for their particular meanings-in-action. These norms are local in that they are distinctive to a particular set of individuals who come to share certain specific local understandings and traditions. Examples of norms developed in the school classroom might be that the use of assessment to inform instruction is central to students' classroom performances.

Finally, the assessment-curriculum relationships and its consequences studied in this inquiry are seen to vary according to the particular circumstances of specific teachers, with specific sets of students, in specific
schools. Thus, their identities are constructed and modified through subjective, participative understanding derived from interaction with their local environments.

3. Delimitations and Limitations. This study was delimited to the third, and sixth grade Language Arts classes, teachers, students, administrators and parents associated with each of two schools. In addition, board-level coordinators restricted to the Benchmark Language Arts development were consulted. While the classes and participants in one of these schools were operating under a developed (i.e., implemented) Toronto Benchmark Program, those in the other were in the early stages of developing (i.e., not fully implemented) their Benchmark Program. While the Montrose site was in its fourth year with the Benchmark Program, the Arcadia site was in its second year. In turn, the clarity, scope and depth of the data from participants were influenced by their time with the Program. The major sources of the data collection were the (participant) observations and testimonies of teachers, students, administrators, and parents associated with each school setting. The (semi-structured) interviews were complemented by the analysis of Board and teacher documents. The data-collection fieldwork process took six months, and therefore gave a "snapshot" of the ongoing Benchmark Program implementation process.

Each of the delimitations carried its own limitation on the study. The data in this study was restricted to Board and school documents, interview transcripts, teachers' diaries, and the researcher's field notes. Although the data-collection period was relatively short, interviews with all participants were detailed and in-depth. The critical-case sample is justified, since those involved were the most representative in being able to report the assessments
and curricula being investigated. However, their involvement means that caution was needed in interpreting statements about positive and negative outcomes, and about difficulties associated with their recall of past events.

4. Significance of the Study. This inquiry examined the perceived consequences of assessment on the curriculum as it related to classroom instruction, and school practice. Examination of the practices and consequences of the assessment-curriculum relationship within the context of educational change presents potential benefits for students (learning more focused on the expected and taught curriculum); for teachers (instruction informed by assessment); for administrators (improved accountability for teacher instruction and student assessment based on school goals); and for parents (explicit and continuous communication regarding the intents, processes and results of their children's school performances).

This research adds to the sparse literature and knowledge base concerning the way in which the assessment-curriculum relationship in the developed, and in the early stages of implementation of the Toronto Benchmark Program affected instructional and assessment practices. It addresses how the diversities in student abilities, socio-economic backgrounds, and interests can be informed by classroom assessment data. Through detailed and comprehensive description of the contextual setting, this study enables the reader to make informed judgments regarding the appropriateness of utilizing its findings and/or framework in other settings.

Further, this study is timely. Its findings helped to provide an empirical basis for critiquing assumptions that either assessment-dominant or curriculum-dominant curricula will impact positively on classroom activities. The findings may have implications relevant to improving students' classroom
performance through a particular mix of assessment and curriculum best suited to one's own educational context.

5. Terminology. Educational change is referred to as a fairly significant shift from one set of attitudes, beliefs and actions to those of another, such as from teacher-centred to student-centred instructional practices in an innovative assessment program. In this study, change implies "planned" change - a non-neutral term implying at least a preferred direction, if not a valued outcome (Leithwood, 1986).

The innovation in this study is the Toronto Benchmark Program, an objectives model of instruction based on achievement-level standards that emphasize mastery learning. The 'benchmarks' are normative standards that describe expected performances on critical learning outcomes at key stages in students' education associated with the grades 3 and 6 Language Arts programs in the Toronto Board. As an innovation, the Toronto Benchmark Program entails significant departures from standard school practices. Fullan (1982) explains that innovation may represent a dynamic interaction among three dimensions: 1) the use of new materials; 2) new teaching approaches; and 3) the alteration of beliefs associated with a new program.

Thus, the term "program" and "curriculum" are used synonymously in this study. Either term is used to refer to organized learning containing a statement of aims and specific objectives. Either indicates some selection and organization of content, patterns of learning and teaching, and includes the evaluation of outcomes (Taba, 1962).

It is important that the development of curriculum not be confused with its implementation. The former refers to the preparation of curriculum plans and materials (i.e., antecedents) for use by students or teachers prior to using
them in the classroom. The development stage involves "getting the curriculum into shape". The latter refers, in part, to "...the process of putting into practice the essential characteristics of an innovation" (Fullan, 1975, p. 1). Additionally, implementation is seen as more subtle than mere fidelity to some specific blueprint for reaching a set of educational goals. It is also seen as a process of mutual adaptation (i.e., curriculum transactions) in which the innovative project and institutional setting adapt to each other (Berman & McLaughlin, 1975). The operationalizing of the curriculum (i.e., transactions) include the teacher-student exchanges, and classroom activities in the Benchmark innovation.

Correspondingly, it is important that assessment and evaluation activities associated with the term "curriculum" not be confused with each other. Assessment is used to refer to collecting data that describe student performances - activities that routinely occur in teachers' classrooms, and which have direct impact for students in their classrooms (Earl & Cousins, 1995). They include such practices as oral questioning, teacher-made tests, portfolios, assignments, and observations of performance. These practices may be administered formally or informally to students, and form the basis for many of the educational decisions that are made by teachers, students, administrators and parents. Performance assessment refers to the observation and rating of student behavior and outcomes in a context where students actually demonstrate proficiency (Stiggins & Bridgeford, 1985). By extension, evaluation is used to refer to the process of making judgments based on the collection of data on student performance. These judgments may be either "low stakes" or "high stakes". Low-stakes evaluations (e.g., MET grade 3 tests used to diagnose learning) are those whose results are seen as not having important rewards or sanctions tied directly to test performance. High
stakes (e.g., the public ranking of schools' test results as an indicator perceived to be important in education), on the other hand, are seen to have important sanctions or rewards associated with test performance. As well, where low-stakes testing is often school controlled and used for instructional purposes, high-stakes testing is often externally controlled and used for measurement purposes (Madaus & Kellaghan, 1992). However, the Province of Ontario, home of the two school sites in this study, is unusual in that it has tests externally mandated but of low consequence for students and teachers.

Frequently referred to in this study is the relationship between assessment and curriculum. The two are viewed as having reciprocal influences on each other. Where the relationship is curriculum dominant, the assessment activities are seen to be subordinate to, controlled by, the tenets of the curriculum. Alternatively, where the relationship is assessment-dominant, the curriculum is viewed as subordinate to the practices and data generated by the assessments. For example, where it is known that specific content will help to make up the substance of a test, that content will likely be taught in class ahead of content that may not be tested. The issue, immediately, is significant to student performance, and to perspective in the selection of actual course content since the time allotted for instruction is limited to that specified in the course timetable. Because time is a finite variable in this study, an emphasis in one area of course content often means a de-emphasis in another area of course content.
A review of the research related to this study is presented within the context of educational consequences evolving from the emphases placed on assessment and/or curriculum. The articles for review in this chapter were located through: 1) a computer search of the Educational Research Information Centre (ERIC) database conducted for the period 1984 to 1997, 2) research reviews, 3) references in articles relevant to the question of this study, 4) my personal source of materials related to assessment and curriculum, and 5) communication with researchers working in this area.

1. Educational Change. Throughout the literature, the terms "change" and "improvement", along with "process" and "difficulty" are emphasized (Berman & McLaughlin, 1976; Fullan, 1991; Huberman & Miles, 1984). For example,

...change is viewed as a process rather than as an event. Change is not automatically accomplished...by a memo decreeing that change will occur...[It] entails an unfolding of experience and a gradual development of skill and sophistication in use of an innovation; it is a developmental process which takes time (Hall, 1979, p. 204).

Rudduck (1991), in referring to the irony of effecting change, comments that we do not have the luxury of closing down our schools in order to desocialize our educators, make changes to the physical site, and get widespread support from pupils and parents. Teachers have to maintain their commitment to the
status quo while planning new approaches; they have to fulfill present
expectations at the same time they consider strategies for changing those
expectations. The appropriate virtues for change would seem to be
imagination, patience, and immense fortitude.

Part of the difficulty in understanding the educational change process is
attributable to the competing conceptions of change as viewed by researchers.
Researchers Snyder, Bolin and Zumwalt (1992) provide an historical
overview of the approaches to curriculum implementation. In the fidelity
perspective, Fullan and Pomfret (1975) viewed the outcome of curricular
change in terms of program materials, teacher beliefs and instructional
practices that are in compliance with the intentions of an original plan, i.e., the
exact "installation" or "replication" of a proposed curriculum. Second, in the
mutual adaptation perspective, Berman and McLaughlin (1976) said that
"...the issue of implementation is often more subtle and complicated than mere
fidelity to some specific blueprint for reaching a set of educational goals" (p.
349). They went on to say that "...local school systems are so structured that,
in order to implement significant innovations, there must be a process of
mutual adaptation " (p. 349). The initial design of an innovative project must
be adapted to the particular organizational setting of the school, classroom and
at the same time, the organization and its members must adapt to the demands
of the project. Each stage in their change process - development,
implementation, and incorporation - was viewed in terms of the interplay
between the project's characteristics and the local setting. Accordingly,
negotiation and flexibility are required of participants adopting this change
perspective. Third, from the curriculum enactment perspective, change was
viewed as the educational experiences jointly created by student and teacher.
The externally created curricular materials and programmed instructional
strategies were seen as tools for students and teacher to use as they construct the enacted experience of the classroom. Thus, the form and function of an enacted instructional program evolve from the experiences uniquely and closely related to the interactions between the participants involved. Fourth, the school-based curriculum development (SBCD) is a form of decentralizing curriculum development to the local school level (Brady, 1990). It is relevant to this study in that it intends to provide a forum for teachers and administrators to assume the central role in developing curricula (Pedretti, 1994), and in promoting ownership, understanding, and acceptance of their constructed curriculum within the terms of their local context.

The comparison of elements particular to the fidelity, mutual-adaptation, enactment, and school-based curriculum models points to the intent of each model. It can be seen that a continuum exists amongst these models in terms of the emphasis in allowing context to be a part of the change process. The fidelity model emphasizes implementation: either the local context can be fitted into the implementation or it is re-shaped. The enactment model emphasizes development which integrates the local context in the change process while the mutual adaptation model seeks a balance between local context and implementation. Furthest from the values of the fidelity model, the school-based model focuses on the culture within the school organization to bring about change. As a cautionary reminder, the research confirms that no single method [of implementation] will permit a universal implementation of new programs. Rather, the research proposes that a variety of methods and assumptions need to be incorporated into any program shift.

Neale, Bailey and Ross (1981), in an analysis of over 25 descriptive theories of the organizational change process, concluded that all models depict three stages of change: adoption, implementation, and continuation. A fourth
stage, outcome, is also pertinent to this study. Variously labelled adoption, initiation, installation or mobilization, development (the term used in this study) consists of planning for change, i.e., the process which leads up to and includes a decision to adopt or proceed with an innovation. Implementation, or initial use, involves the first experiences in attempting to put an idea or program into practice (Fullan, 1975); and in adapting the innovation to the environment and the environment to the innovation (Berman & McLaughlin, 1974). Continuation refers to whether the change gets built in as an ongoing part of the system or disappears by way of a decision to discard or through attrition (Berman & McLaughlin, 1978; Rosenblum & Louis, 1979; Yin, 1984). "Outcome" was added to the process by Fullan (1982), and speaks to the degree of school improvement in relation to given criteria. These components suggest that change may be more or less defined in the early stages (i.e., development), and move to attempted use (implementation). If the use is accomplished, it may be sustained over a longer term (continuation). From the extended use, different types of results (outcomes) may evolve.

Research has shown that numerous factors may influence the process of change in each phase. Most theories have conceptualized initiation as a set of actions that centre on the official adoption of a prepackaged innovation in its complete, and intended state. The practices recommended by the innovations were assumed to be superior to existing practices, and sometimes included the process by which this technology was to be put into practice. Variations in sets of actions can be accounted for by the differences in perspectives from which each views adoption. These concepts can be classified as problem solving at a point in time (Gross et al., 1971; Giaquinta, 1973; Zaltman, 1973); as activities designed to attract consumers and train users following the invention and development of an innovation (Carlson, 1965; Berman &
McLaughlin, 1975; Hall & Loucks, 1978); and as administrative declarations of adoption and the allocation of organizational resources (Charters & Jones, 1973; McKinney & Westbury, 1975). Berman (1981) analyzed development into four sub processes: policy image developments; planning; internal support generation; and external support generation. Obviously, development has been viewed as the beginning of the importing of the innovation into the system by school authorities despite the minor variations in perspective.

Table 1 outlines those factors found in the research (Fullan, 1982) critically related to the first stage of curriculum change:

<table>
<thead>
<tr>
<th>Factors Associated with Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Existence and quality of innovations.</td>
</tr>
<tr>
<td>2. Access to information.</td>
</tr>
<tr>
<td>3. Advocacy from central administration.</td>
</tr>
<tr>
<td>4. Teacher pressure/support.</td>
</tr>
<tr>
<td>5. Consultants and change agents.</td>
</tr>
<tr>
<td>6. Community pressure/support/apathy/opposition</td>
</tr>
<tr>
<td>7. Availability of federal or other funds.</td>
</tr>
<tr>
<td>8. New central legislation or policy.</td>
</tr>
<tr>
<td>10. Bureaucratic incentives for adoption. (p. 42)</td>
</tr>
</tbody>
</table>

The ten factors, depending on their presence or absence, influence or result in decisions to reject or develop specific change programs, policies or directions. As well, the Eight Year Study by Smith & Tyler in 1942 revealed determinants later expressed by teachers (Table 2) as critical to the development of their curriculum. Teachers and administrators involved in the
30 schools in the study reported they needed the following 'determinants' in order to develop a curriculum successfully:

Table 2

<table>
<thead>
<tr>
<th>Determinants of Development from the Eight Year Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Time - especially for teachers to study and plan together;</td>
</tr>
<tr>
<td>2. As broad a base of participation as possible - including parents and students;</td>
</tr>
<tr>
<td>3. Cooperative coordination (comprehensiveness);</td>
</tr>
<tr>
<td>4. Research - both general findings and specific local conditions;</td>
</tr>
<tr>
<td>5. Planning - It was the nature rather than the amount of planning talk that was problematic. In retrospect, many teachers felt they talked too much, but blamed it on their ability to work in a group and lack of experience in &quot;real&quot; thinking. &quot;The schools that plunged into change without taking time to think their problems through often found it necessary to go back to the beginning and start over&quot; (Aiken, 1942, p. 127).</td>
</tr>
<tr>
<td>6. Continual internal evaluation - to record, report, and analyze what took place:</td>
</tr>
<tr>
<td>7. Willingness to experiment and change:</td>
</tr>
<tr>
<td>8. Conviction and courage;</td>
</tr>
<tr>
<td>9. Collaborative leadership and teacher collaboration. Without teacher ability and willingness to collaborate, the enactment process could not be fulfilled; and</td>
</tr>
<tr>
<td>10. Freedom and responsibility (Giles, McCutchen and Zechiel, 1942).</td>
</tr>
</tbody>
</table>

Researchers (Aiken, 1942; Giles, McCutchen & Zechiel, 1942) comment that change is not merely observable alterations in behavior, but rather a personal and continual development process, both for the teacher and the student. Successful developmental change process requires the understanding and acceptance of the subjective realities of the participants involved. In turn, they believe the role of the teacher to be that of curriculum developer who, together with his or her students, develops positive educational experiences.

The determinants from Tables 1 and 2 show that innovations may be developed from many different sources, and for different reasons. The need
for change may be embedded in any one or several of the factors. Their order is not important, but their combinations are. For example, teacher and parental access to an innovation that is clearly conceptualized may elicit a different kind of support than would a less clearly thought-out innovation.

In Table 3, fourteen critical factors organized into four main categories relating to 1) the characteristics of the innovation; 2) the strategies; 3) characteristics of the adopting unit; and 4) the macro socio-political units that summarize the influences impinging upon implementation (Fullan & Pomfret, 1975).

Table 3

<table>
<thead>
<tr>
<th>Variables Influencing Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Characteristics of the Innovation:</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>B. Strategies</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>C. Characteristics of the Adopting Unit</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>D. Characteristics of Macro Socio-political Units</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
The factors in Table 3 represent a system of variables which interact with each other. Certain determinants might be more pertinent to the implementation than others. For example, the determinants of plans for explicitness and complexity or degree and difficulty of change required by the innovation are critical. Low innovation explicitness is usually associated with user confusion, lack of clarity, frustration, and low degree of implementation. Also, the more difficult the change, the greater the degree of learning entailed by it, and the more likely that the degree of implementation will vary across groups of users. Determinants pertinent to the implementation phase will be those that support the basic purpose of fidelity implementation: the resocialization (i.e., of internal values, external teaching behaviors) of key participants. Thus, the determinants of inservice training, resource support, feedback mechanisms, and participation in decision making are critical to the change process. Similarly, as the organizational system and incentives would be critical to the continuation process, evaluation would be critical to the outcomes phase of implementation.

From an analysis of 15 research studies on curriculum and instruction, Fullan and Pomfret (1977) identified five dimensions of implementation in practice: "changes in a) subject matter of materials, b) organizational structure, c) role and behavior, d) knowledge and understanding, and e) value internalization" (p. 361).

In the Rand Change Agent Study, Berman and McLaughlin (1976) investigated the process of innovation and the factors affecting it through a [U.S.A.] nationwide survey of 293 federally-funded change agent projects. The researchers examined a variety of innovations (classroom organization, bilingual programs, reading projects) using case studies (N=29), and
questionnaires with the participants. Table 4 (Berman et al., 1976) summarizes the factors that are seen to have affected the implementation and continuation of innovations in the Rand Change Agent Study.

Table 4  
Factors Affecting Implementation and Continuation

A. Project Characteristics
1. Type of educational treatment
2. Resource level
3. Substance and scope of proposed change
   - perceived educational priority
   - requirement for change in teacher behavior
   - comprehensiveness and complexity of required changes for a number of people
   - consonance between the values and goals of the project and those of the staff
4. Implementation strategy
   - on-line planning
   - practical staff training
   - local development of teaching materials
   - 'critical mass' of staff working on project

B. Institutional Setting
1. Support from the school principal and district administrators
2. High teacher morale
3. Teacher willingness to make an extra effort

C. Federal Policies
1. Implications for the initiation stage

The researchers found that the purpose common to all the change agent programs studied in the Rand Change Agent Study was "...the stimulation and spread of educational innovations" (p. 346). They developed three obvious measures of the effectiveness of an innovative project's implementation:
perceived success, change in behavior, and fidelity of implementation (p. 350). While the Rand study on innovative programs is the most comprehensive study of a large number of educational innovations researched to date, serious concerns exist regarding their measures of implementation. All five of their measures were global in nature and do not require specific knowledge of the dimensions of implementation. Faith in the measures of implementation is further weakened by the low correlations of their first four measures - nearly all of which were below .35 (Berman & Paully, 1975, p. 17). Nevertheless, the Rand Change Agent Study is still well respected for its findings on determinants of implementation.

The research on SBCD emphasizes the problems associated with it: the lack of individual schools to manage curriculum change, and the lack of capacity to track the assessment of the proposed changes (Fullan, 1992). Levine and Eubanks (1989) research contains an excellent analysis of the problems of local school reform that uses school-based models and empowerment assumptions. They identify six obstacles: 1). inadequate time, training and technical assistance; 2). difficulties of stimulating consideration and adaptation of inconvenient changes; 3). unresolved issues involving administrative leadership and enhanced power; 4). constraints on teacher participation in decision-making; 5). reluctance of administrators to give up traditional prerogatives; and 6). restrictions imposed by school boards (pp. 117-118). In citing several examples, Levine and Eubanks also report that school-based management projects encounter problems over delegation, training and skill requirements, and taking action to accomplish the projects' goals. Other researchers (David, 1989; Ogawa and Malen, 1989) claimed that shared governance [in the eight schools they studied] had not only failed to alter traditional decision making relationships, but they also served to reaffirm
them by defusing important issues and by developing loyalties without targeted action. Levine and Eubanks caution us of three dangers with SBCD: 1). that satisfaction with the SBCD process is often confused with actual curriculum gains; 2). that the responsibilities for failure as well as success in implementing the innovation might be shifted to the local school participants; and 3). that SBCD has not yet demonstrated that it focuses on, let alone alters, the changes required for reform.

In spite of the problems associated with SBCD cited, research (e.g., Elmore, 1990; Fullan, Bennett & Rolheiser-Bennett, 1990; West & Hopkins, 1996) continues to seek simpler and more workable components to restructuring schools by adopting a broader view of student performance and the instruction which might help to develop it. For example, West & Hopkins (1996) are attempting to test a comprehensive model for school improvement that seeks to integrate 1). student experiences; 2). student achievements, 3). teacher and school development, and 4). community involvement. Much of the current research attempts to integrate approaches from school effectiveness research and school improvement practice. The goals of this research include finding guidelines to support collaborative school development, and establishing sound educational principles upon which to base school improvement.

Two conclusions that may be drawn from the research on SBCD. First, sustained improvement requires restructuring of the school, the Board, and their interrelationships. The roles of students, teachers, principals, and parents are implicated as is the structure, governance and design of the reform (Elmore, 1990; Murphy & Evertson, 1990). Second, it is very difficult, if not impossible in many cases, for schools to redesign themselves. As a result, the
role of the Board is crucial in providing support to ensure the improvement and continuance of school-based reform.

The review of the literature dealing with educational change addressed curricular consequences at the classroom and school level. This section on the change literature shows that while conditions that may influence each phase of the change process have been identified, the linkages between each phase of change have not. Thus, this study is well-focussed on the relationships among the antecedents, transactions, and outcomes of the innovative Toronto Benchmark Program.

2. Classroom Assessment. The following section of this chapter presents literature particular to the meanings and assumptions attached to traditional and performance assessments.

i. Traditional. Traditional assessment is often referred to as 'objective' testing. Traditional tests are typified by the standardized, centrally-developed and scored provincial examinations, multiple-choice, fill-in-the-blanks, end-of-course exams, pencil-and-paper, and short answer testing formats. Their results can easily be reported as a numerical score, i.e., a median score used to locate a student's "rank" relative to a larger population.

According to Tyler (1949), in his often referenced Basic Principles of Curriculum and Instruction, the process of assessment begins with the objectives of the educational program. Since the purpose is to see how far these objectives are actually being realized, it is necessary to have assessment procedures that will give evidence about the behavior implied by the objectives (p. 110). However, the view that objectives should be stated clearly in terms of students' behavior was common to industry during the early part of this century. For example, both Bobbit (1918) and Charters (1924) were
instrumental in developing the theory and practice of objectives in our schools. By 1949, the objectives approach was firmly established. More recently, the growing acceptability of the objectives approach can be seen with the introductions of the 'common curricula' in Great Britain, and Canada (Ontario) as well as in the United States in conjunction with performance-level standards. Underlying the objectives approach is an assumption that schools can function effectively only if they have adopted clear aims and objectives with ways of showing whether or not they have been achieved.

Tyler helped to promote the objectives approach as director of evaluation of the Eight Year Study between 1934 and 1942. He saw his role as facilitating teachers in the 30 participating schools to formulate educational objectives and then to develop assessment techniques to measure them. Tyler's study has particular relevance for this research in that it found evaluative practice as following the objectives of the curriculum while at times it was the tests themselves that came to define the curriculum and its objectives (Travers, 1983). Tyler commented that in selecting objectives for classroom instruction, care needs to be taken to ensure that the objectives chosen are significantly relevant to the content of instruction; that the objectives are consistent with each other, and that they are attainable by the students. Tyler believed that because the purpose of education is to bring about significant changes in student behavior, the objectives must be statements of learning so that the appropriate instructional activities can be planned.

According to Tyler, the evaluation process helps to provide data on student and teacher activity that helps to identify the strengths and weaknesses of the curriculum program, in turn, helping to clarify the objectives themselves; provide feedback on the appropriateness of the learning experiences in fulfilling these objectives; identify student weaknesses and strengths for
remediation and enrichment; establish a basis for planning programs appropriate to the needs of students; and provide a basis for communicating the school's success to its wider community.

Influenced by Thorndike's (1918) belief that education was concerned with changes in human beings and that 'changes' could be known only through what people "do", schools began to use objective achievement tests across the curriculum. Valuing behaviorism, these tests assessed the efficiency of teachers and school systems in developing curricula during the 1920's, while in the 1930's they were used almost exclusively to evaluate individual students - to assign grades, to diagnose learning difficulties, and to place students in instructional groups (Cronbach, 1983). After World War II, the development and use of standardized tests continued to grow, behavioral objectives (Tyler, 1949) emerged, and a classification to help educators articulate levels of student learning was developed (Bloom, 1956).

ii. Performance. Currently, there is a strong movement toward using performance assessments to supplement or even replace traditional testing (Hill et al., 1996; Lewis, 1996; Messick, 1994; OECD, 1993). Performance assessments may be found in such formats as classroom projects, lab experiments, demonstrations, and dramatizations. Students are asked to construct their own understanding and demonstrate mastery in their own ways on concrete, well-contextualized tasks or activities that are sampled from a subject-domain. Usually, a performance assessment is scored by combining both the conclusions drawn from an activity and the procedures used to arrive at those conclusions.
For the purpose of this study, the description of performance assessment given by Stiggins (1994) is used:

Performance assessments involve students in activities that require demonstration of certain skills and/or the creation of specified products. As a result, this assessment methodology permits us to tap many of the complex educational outcomes we value that cannot be translated into paper and pencil tests. With performance assessments we observe students while they are performing or we examine the products created, and we judge the level of proficiency demonstrated. (p. 160)

Performance assessments may be based on observations of the process while skills are being demonstrated, or on the evaluation of products being created. Evidence of achievement is in 'the doing' and/or in 'the product'. The index of achievement typically is a performance rating or rubric that reflects the levels of quality in the performance. Stiggins (1987, 1994) agrees that the purpose of performance assessment is to assess a student's ability to translate knowledge and understanding into action, and that the student's response is to plan, construct and deliver an original response. He believes the beneficial attribute of performance assessment is that it helps us see clearly the evidence of performance skills, and that this evidence helps to inform students' learning.

Performance assessment is grounded in assumptions relating directly to setting standards, internalizing and achieving standards, and student performances. First, we may set public educational standards and strive toward them. These standards for student performances (performance criteria) are the specific requirements of performances, including the knowledge, concepts, skills, and processes that might be exemplified in classroom performance (Stiggins, 1988). Levels of performance quality
(performance standards), usually ranging from novice to expert, accompany the performance criteria (Wiggins, 1990). Examples of student work, i.e., exemplars, are then obtained to make these statements of expectations concrete and tangible. Second, researchers see performance assessments as central to the instructional process (Resnick & Resnick, 1991; Wiggins, 1989). A student's work either does or does not achieve the standards for that type of performance. Students may need varying amounts of time and instructional methods, but the goal is the same for all: achievement of the standards (Bloom et al., 1981). Third, very different student performances and exhibitions may reflect the same standards, i.e., it is likely that a range of performances (pertinent to the characteristics of the domain tested) based on each standard can exist. Fourth, educators may be trained to internalize the standards, and judge student performances fairly. For example, teachers can assess depth and accuracy in student writing through understanding the attributes of good writing (Stiggins, 1992).

3. Consequences of Classroom Assessment. The positive as well as negative consequences for traditional and performance assessments gleaned from the research on assessment and curriculum are presented in this section.

i. Traditional. The most direct measures of schooling outcomes are obtained from assessments of students' academic performances at the classroom level. Herein, however, lies a problem that has been evidenced through disciplined critiques of traditional [and prevailing] psychometric models for test and examination modes of assessment (e.g., Berlak, 1992) and an equally voiced concern for the negative effects of test-driven and test-dominated curricula. Frederiksen (1984) expressed this concern aptly: 'the tail of testing that wags the curriculum dog' (p. 201). To date, the
measurement of students' learning outcomes at the classroom, school, board, national and international levels has relied almost exclusively on the use of standardized achievement tests or external examinations (Haney, Madaus & Lyons, 1993; Madaus & Raczek, 1996; Goldstein, 1996; Goldstein & Lewis, 1996; Sutherland, 1996). Although the use of traditional tests and examinations for the measurement of student performance is usually justified on the grounds of maximizing reliability and ensuring comparability, it is argued that this has been achieved at the expense of validity (Broadfoot, 1994; Moss, 1994).

In both the US and Great Britain, there has been mounting criticism of the utility of standardized tests as measures of either learning or competence (Broadfoot, 1996; Gipps & Murphy, 1994; Murphy & Broadfoot, 1995; Newmann & Archibald, 1992; Wolf, 1995). For example, Newmann and Archibald (1992) argue, that "most data currently used to assess scores on standardized tests fail to measure meaningful forms of human competence and that significantly new forms of assessment need to be developed" (p. 164) while Broadfoot (1994, p. 5) comments that "dissatisfaction with traditional testing approaches is now widespread, explicit and clearly articulated". Elements of this criticism have been applied to the areas of standards monitoring and performance assessment in North America where alternative approaches to obtaining more curriculum-specific and authentic (Wiggins, 1989b) measures are being implemented (Floden, 1994; Shavelson, 1994; Taylor, 1994; Nitko, 1995b). Similarly, the research inquiries of Butterfield (1995), Gipps (1994a, b, c), Gipps and Murphy (1994), Torrance and Pryor (1995) and Wolf (1995) are representative of a educational researchers who are calling into question prevailing "mechanistic", "objectives-driven" modes of assessment and are suggesting that the professional role of the teacher be
re-established by challenging the widespread assumption that teachers' assessments are less reliable than those obtained from examinations and tests.

Specifically, research has shown that traditional (i.e., standardized, objective) assessment has the capacity to shape the curriculum (Ascher, 1990; Burstein, 1991; Neill, 1991; Shepard, 1990). Undeniably, curricula and school reform are affected by the tests and their scores. Tests which are centrally controlled may undermine such reforms as school-based, and shared decision-making, thereby leaving educators accountable but powerless. A dual curriculum may exist as an outgrowth of traditional testing. One curriculum may focus on an outcome-based curriculum, often called performance-based or authentic assessment; a second curriculum may focus on concepts in preparation for the test-taking (Ascher, 1990). Although much of the public favors high-stakes testing (Ascher, 1990), such testing may pose hurdles to students who are most in need of educational support. Teachers tend to teach to the test and a 'de facto' national curriculum based on test material tends to emerge (Brophy, 1990; Kilian, 1992). This process can eliminate or reduce instruction in areas not covered by traditional tests. Further, students cannot receive assistance on traditional standardized tests; such tests usually do not have strong instruction-like qualities (Lowenthal, 1988).

Nagy & Traub (1986) identified four effects of standardized assessments on the teaching-learning process. First, each teacher could evaluate his own teaching objectives and content coverage by examining the published samples of provincial assessment items and objectives. In the case of the 1991 Alberta Achievement Testing program, sample questions and test information were given to all participating schools prior to testing. The sample questions and test information are able to inform and guide the levels of teacher instruction, and student learning outcomes. Second, as a group, teachers might use
assessment results to assess how effective their teaching strategies have been in relation to student performance on particular assessment items and objectives. Third, in addressing exceptional student responses to assessment items, teachers might try to suggest new strategies in order to produce better results. Fourth, focus is directed toward textbook and other instructional aids in order to improve student performance on curriculum objectives deemed to be important.

Worthen and Spandel (1991) discuss other common consequences of traditional tests. Such tests are relied upon by schools to accommodate organizational needs of accountability and may not directly promote student learning. Standardized tests are poor predictors of individual student performance. Next, the content of standardized tests is often mismatched with the content emphasized in a school's curriculum and classrooms. The curriculum is sometimes narrowed down by teachers in order to emulate the curriculum suggested by the test, and neglect other important areas. Students are often categorized, and labelled according to their scores on standardized tests. As well, standardized measures are suspect in being racially, culturally and socially biased. Last, these researchers state that standardized tests are criticized for measuring only declarative (i.e., limited and superficial) student knowledge rather than more procedural thinking skills.

Other research findings on the consequences of traditional assessments include the 'dumbing down' of instruction - teaching the precise content of the tests rather than the underlying concepts (Wise, 1985); 'entrenching' of the curriculum, i.e., placing all the onus on the student to adapt to the instruction and learning (Gickling & Thompson, 1985); and reducing sensitivity to measuring change in student achievement (Marston & Magnusson, 1985).
Others (Hambleton and Murphy, 1992; Worthen & Spandel, 1991) find that objective tests foster a 'one-right-answer' mentality.

However, some researchers argue that the evidence against multiple choice tests is not as strong as has been claimed (Worthen & Spandel, 1991), and that more research into the strengths and weaknesses of traditional assessment formats for meeting particular measurement needs should be carried out. Rather than eliminate the use of traditional assessment, the researchers suggest that they supplement or complement performance assessments in our classrooms. While there has been extensive research for over 25 years on the impact of standardized testing on students, there has been little research until recently aimed at the kind of assessment that routinely takes place in schools and classrooms; assessments that can form the basis for decisions to be used to promote the development of student performance.

ii. Performance. Gamlin (1989) contends that performance assessments take into account students' prior learning; attend to instruction, e.g., by helping educators gear programs to student needs; see student abilities as modifiable through direct instruction and assessment; seek a 'match' between instruction, and student responsiveness to new material; ensure that assessment results are updated continually; and promote student equity (Darling-Hammond, 1994). When multiple performance assessments are utilized (Taylor, 1989), the results are likely to be relevant to each student's learning and instruction, since the test items given to the student are often drawn directly from the intended curriculum (Bachor and Crealock, 1986). The teacher is then able to make informed decisions concerning what and how students need to be taught. Depending on their specific needs, students might be asked to demonstrate their knowledge through various modes and levels of
complexity, and so generate data enabling teachers to individualize their instruction. In a review of the literature, researchers (Haydel, Oescher & Banbury, 1995) comment that performance assessments reach quality status when the purpose of the assessment is clearly articulated, when the target is articulated and focussed, and when the assessment is closely matched to the assessment target. Second, quality performance assessment arises out of articulation of performance criteria, i.e., the performance criteria are specified; the performance criteria are expressed in terms of observable behaviors or products; the criteria are comprehensive, reflecting the essential components of the task; the criteria are developmentally appropriate for the student; and the performance assessments are comprehensible to both the teacher and the student. Third, quality performance assessment arises out of a consideration of the setting in which the assessment occurs, i.e., the student performance relative to the performance criteria can be demonstrated in the setting in which the assessment takes place; and the student performance can be assessed in the setting. Fourth, quality performance assessment arises out of the scale by which performance criteria are scored. The scoring scale should represent an underlying continuum of quality relevant to the performance criteria; points on the continuum are specified; and these points differentiate the quality of the performance. Fifth, maintaining a written record of student performance and managing the results are essential to quality performance assessment. The scoring record should document the performance of students on the established performance criteria; summarize the assessment using the collected data; and communicate the results of the assessment. Sixth, quality of performance assessment arises out of its format. To accomplish this, the entire performance assessment must be organized; and it must use standard
writing conventions (i.e., syntax, usage, capitalization, punctuation, and spelling).

As with any other measurement technique, performance assessments also have their disadvantages (Mehrens, 1992). Haydel et al. (1995) found that teachers' performance assessments may not be as sound as they could be due to shortcomings in defining the purpose of the assessments; matching the method of assessment to the target; and articulating the performance criteria, the scoring scale, and the scoring record. With only a limited number of assessment tasks, it is difficult to keep the exact content of the exam secure. On the other hand, Popham (1994) criticizes some performance assessments for not adopting test-item specifications that have only single definition responses: with multiple definition responses to each item, tests can quickly become too unwieldy to administer and score. Popham (1994) suggests that we must try to:

...understand the essence of the intellectual operations that the student must perform in order to complete the tasks represented in the test's items. What we need to focus on are the various kinds of test items that will help us make accurate inferences about a particular kind of ability. (p. 17)

However, Nitko (1995b) explains that developing curriculum-driven assessment is a very labor intensive process, and doubts that schools alone would have either the people or the time to develop such programs. Instead, he suggests that the development of CDA programs be left to the more resource-rich [provincial] levels of government. The literature also identifies that once performance assessments have been used, they cannot be re-used to test the same higher-order thinking process. One can memorize a sophisticated answer as well as one that is very basic. Thus, performance
assessments would have to be modified each year - adding to the developmental costs, and difficulties in making cross-year comparisons. Frechtling (1991) adds that performance assessments can be very time-consuming - they may extend over days or even weeks; tend to provide a multi-dimensional look at a particular skill or area - so that breadth of coverage may have to be traded off for depth of coverage; take considerable teacher time to administer - making it difficult to provide 'make-ups' for students who are absent or who have to be retested with an alternative form of the same test; involve a frequently complex scoring procedure - as a result, the end product may be a series of scores or a narrative summary of skill attainments for each assessment rather than a single figure; and frequently include the classroom teacher as a critical factor in the scoring process - introducing more subjectivity into the scoring process.

The attacks on performance assessments often come from the proponents of externally mandated tests who are very occupied with psychometric concerns, e.g., issues relating to reliability (Linn, 1994). Studies on performance assessments tend to find inconsistent results regarding the reliabilities of these tests. For example, while Le Mahieu, Gitomer and Eresh (1995) found high inter-rater reliabilities for writing assessments, other research (reviewed in Bateson, 1994) did not. In addition, Koretz, Stecher, Klein and McCaffrey (1994) found little consistency between the scores students received on different writing samples (even though the samples were intended to represent the same tasks). Nevertheless, it should be pointed out that the correlations for the videotaped and written tasks in the lessons of the Toronto Benchmark Program in this study varied from .7 to .96 (averaging in the mid .7's), and from .79 to .94 (averaging in the mid .8's) respectively (Larter, 1991). Second, in the context of assessment for instruction and from the perspective.
of the classroom teacher, reliability may be considered a minor technicality. For example, learning is considered to be contextual, situated (Anderson, Reder & Simon, 1997) within the performance criteria associated with each Language Arts task. In this context, validity is more valued than is reliability since the adequacy and appropriateness of interpretations and actions based on assessment scores (Messick, 1995) is most closely aligned with having students learn particular skills in the classroom.

However, both the traditional and performance assessments discussed in this section are subject to the same criteria of quality control. For example, Stiggins (1994) sets forth a set of guiding principles for high-quality classroom assessment that relate to target, purpose, method, sample, and control of inference. While it is recognized that traditional and performance assessments are aligned, respectively, with the separate functions of measurement and instruction, a central concern in this study is the emphasis placed on the interdependence of assessment and curriculum in which the distinction between assessment and curriculum is removed. i.e., assessment is not an artificially separate but an integral part of the curriculum (Gipps, 1994b; Wood, 1986). The attributes on traditional and performance assessments have been discussed in order to clarify the identity of each, and to clarify the consequences that may evolve from using each assessment format.

4. Research Relevant to this Study. Researchers (e.g., Anderson, 1990; Wideen, 1991) find that the consequences of provincial assessments are consistent with those found in the research literature on this topic. On the positive side, teachers appeared more likely to emphasize the priorities and standards of the provincial curricula. On the negative side, teachers' autonomy and ability to individualize the curriculum were constrained with
instruction narrowed to conform to test requirements. A majority of teachers give acceptable marks to the tests' quality and fairness, but believe they are used more for political than for educational purposes.

In the United Kingdom, research (PACE - Primary Assessment, Curriculum and Experience) on the National Curriculum (Broadfoot, Pollard, Croll, Osborne, & Abbott, 1994) examined the educational consequences on students in primary schools. In the National Curriculum, each subject had prescribed programs of study and standardized assessment with national testing for students at the ages of 7, 11, 14, and 16. The results of the common assessment system were to be made public to parents, and to help make schools more responsive to market forces through the operation of parental choice. The interviews which focussed on classroom practice with regard to assessment, and instruction found that the curriculum was rigid and over-prescriptive, that it stifled teacher creativity, and capacity in responding to the needs of particular pupils and situations. As well, the testing arrangements were criticized for contributing to a sense of failure in some children, and for emphasizing assessment over instruction. For example, one teacher in the study commented:

I am not prepared to become somebody walking round with a check-sheet...I think my place is with the children, making a relationship with them. It's not fiddling around with bits of paper or spending all my time talking with their parents.

The most important area of change involved more time spent on assessment-related record keeping. There were fears that this was beginning to 'take over from teaching', that the heavy burden demanded in time and effort for assessment left too little time for planning, for responding to children, for
display work, and for all the things seen by many as "real teaching". A
second-year teacher reflected on the notion of learning:

You're desperately covering stuff because you must account for it, and you
think, "This is just not what it's about. Learning is not about this, and this
is not what it should be like".

A third-year teacher commented on the need for "right" results:

There certainly are more constraints, more pressure on you to produce results,
having something to show, some evidence that the children have done it.

Another teacher recounted her frustration at being forced to "cover too much
content":

There are often times when an activity is going really well, and I know
they're [students] enjoying it and I'd like to go on and develop it, but I have
to urge them to finish off because there's something else I need them to start.
I feel we're being asked to pack so much into each week that I do think we're
in danger of being shallow.

Still others (one-fifth of the 148 teachers interviewed), saw the National
Curriculum as complementing and enhancing their skills and strengths, or as
providing the opportunity to develop them further. As one teacher stated:

The National Curriculum has been useful in giving us targets and a framework
to work within. In a way, it's a bit reassuring. It gives me a focus and
another reason for doing things. It helps me feel as though I'm on the
right track and releases me to work out new ways of teaching.

These comments by teachers involved in putting the National Curriculum into
classroom use from the Broadfoot et al. (1994) study highlight the
consequences of an assessment-driven curriculum on school and classroom practices.

Torrance (1995) found that teachers' approaches to assessment activities in the (UK) National Curriculum are strongly associated with the quality of learning experienced by the students. For example, convergent assessment (emphasizing whether or not the student knows or can do a predetermined act) is associated with linear instruction and a behaviorist view of learning. Divergent assessment, on the other hand (finding out what the student knows or can do), is associated with individualized instruction and a constructivist view of learning. Torrance's classroom observations reveal that because of constraints, e.g., an over-prescribed and inflexible curriculum, teachers can 'fall into' a behaviorist approach in order to survive the demands of delivering the curriculum content.

Ruddock (1995) highlights a number of concerns with implementing performance assessments based on the UK experience with the National Curriculum. First, a clear purpose is needed for standards. However, the degree of their specificity or generality was seen to be important to how they played themselves out in actual practice. In the UK, Ruddock believes the National Curriculum provided more detail than was useful, and its manageability was impaired as a result, i.e., the relationship between purpose and form needs to be clearer than was the case of the National Curriculum. The manageability of the system needs to be evaluated from the start. For example, an appropriate balance between too many detailed standards and too few over-general standards might be more feasible for classroom use. Second, the standards constructed need to be related to a model in order to be able to make decisions on whether or not the standards have been achieved. Third, the implementation of standards related to assessments needs to be
considered throughout the construction process [of standards] with attention to intended complexity of standards, and matters of context. Fourth, the modes of assessment, and the administration time and arrangements they require, need to be considered throughout the construction process. The scale on which they will need to be carried out also needs to be considered. The author believes that many of the problems experienced in England and Wales originate from this source. Fifth, public understanding, and even effective transmission of the standards, is difficult to obtain if they are too numerous and specific to form part of a complex structure. These problems are accentuated if the system has to be constantly revised. Sixth, standards which attempt to be multi-purpose should be viewed with extreme skepticism. Those which are intended to avoid narrowing the curriculum by including process and practical items need to be evaluated for the manageability of the assessment arrangements if they are to be implemented on a wide scale. Seventh, standards for different subjects need to be coordinated to ensure they do not conflict and, taken together, the assessment time and arrangements are reasonable. Eighth, extensive piloting is desirable in order to assess manageability and to identify aspects of the system which are not of the desired level of difficulty. If the standards are to be accompanied by a new curriculum, rather than superimposed on an existing one, this needs to be allowed for when plans for piloting are drawn up. Ninth, the speed of implementation needs to be controlled to avoid the need for frequent revision to eliminate or reduce the effect of undesirable characteristics. In the UK, this has led to virtually constant change which has helped to demoralize teachers and hamper evaluation. The UK experience identifies areas where the problems with implementing an assessment-curriculum innovation may prove to be difficult to solve in other locations.
Nutthall and Alton-Lee (1993; 1992) conducted observational studies in order to assess the classroom learning of individual students in three separate studies of curriculum units in science and social studies in elementary and middle-school classes. They found that how students fulfill their assessment tasks is at least indirectly determined by the nature of their classroom experiences. The more complex and varied the classroom activities, the more likely they are to produce knowledge structures that are rich in content. The more these students learn in classrooms - the better they may perform in their assessments, not because they have learned more answers, but because they have more avenues for solving the problems posed in their assessments.

Research by Stodolsky and Grossman (1995) on the consequences of subject matter for curriculum found, through survey responses from 399 teachers in 16 schools in California and Michigan, that with more sequential subject matter, teachers have little autonomy in selecting course content, instructional materials, and teaching techniques. Also, teachers were pressed to cover the prescribed content, and were required to administer exams that were standardized in order to standardize their teaching across classes and schools. The researchers noted a tendency for sequential courses to become 'entrenched' through standardization with little change made to their assessment and instruction.

Hertert (1995) identifies state-level reform of an assessment program in Kentucky. Writing and mathematics portfolios were added to performance assessment development in a number of content areas. A continuous assessment instrument (KELP: Kentucky Early Learning Profile) was piloted, and is used to evaluate primary programs (multi-age, non-graded classroom groupings in elementary grades), and to determine a student's readiness for the fourth grade. Correspondingly, the Kentucky department of education has
developed a "school accountability index" with corresponding regulations for schools regarding rewards and sanctions based on students' performance levels. In response, a large number of Kentucky educators felt that it was unreasonable for the state to implement a "high stakes" assessment instrument before developing the corresponding curriculum frameworks. As one Kentucky superintendent stated, "we're being tested and held accountable on a curriculum that hasn't been written...it seems to me that they [the legislature] have some cart-horse issues". Another Kentucky administrator commented "...the state talks about local decision making, talks about flexibility, and then they continue to send program advisors out directly to schools who say, 'you're going to do this'..." In turn, classroom teachers identified time as a critical component in making significant changes in educational practice. Many complained that the state did not provide sufficient time to explore fully, plan, implement, and evaluate the required changes. A Kentucky teacher commented,

...the biggest problem with implementing KERA is time. We need more time for collaborative planning among teachers. In the elementary day, it's just not there. We get 30 minutes per day, including time to grade papers, put up bulletin boards...

Another teacher explained,

Suddenly we are confronted with one-year or two-year timelines in which to change our whole theory, philosophy, and practice.

In Wyoming, four classroom studies (Kleinsasser, Horsch & Wheeler, 1995) examined the expanding roles of teachers involved with performance assessment reform. With 18 teacher-researchers in four Wyoming elementary
through post-secondary school sites, the study set out to describe and analyze the effects of teacher-developed science and math performance assessments on teachers and learners; and to examine performance assessment's potential to expand learner and teacher roles. Overall findings highlighted the value of student self-directed learning and assessment, the value of teacher-student assessment collaboration, and a heightened teacher sensitivity to the instructional process.

Researchers McCallum, Gipps, McAlister & Brown (in Torrance 1995a) documented variations amongst teachers in 32 schools (with case studies in six of the schools) in their perspectives and approaches to assessment. They summarized the variations in terms of three 'idealized types' which they called 'intuitives', 'evidence gatherers', and 'systematic planners'. Only the 'systematic planners' made a close link between assessment and instruction: teachers in this group not only planned assessment activities on a weekly basis, but also used the results in planning instruction, often using the results for diagnostic purposes. These three approaches to assessment were related to the teachers' views of teaching and learning, their general style of organization and teaching, and their reaction to the imposition of an assessment-curriculum system mandated externally (in this case, the National curriculum Assessment in the UK).

These studies all deal with the development, use, and consequences of innovative programs that merge assessment with curriculum. In examining the relationship between assessment and curriculum, these studies highlight the components associated with this research inquiry: standardized testing, outcomes-based education, assessment for measurement, and for instruction, concerns for implementation, and classroom consequences of assessments. In only a few instances have findings been reported through the first-hand voices
of the participants. Where the voices have been heard, the consequences of the assessment-curriculum relationship have been given contextual meaning. On this point, there has not been sufficient research concerning the impact of assessment and curriculum relationships may have on teacher practice, student performance, and classroom culture.

5. **Summary of Research.** This literature review shows that research on student evaluation, and on the implementation of curriculum have, until very recently, been dominated by large-scale testing, and administrative interests. Kuhn's (1982) work with paradigms, i.e., an implicit, unvoiced, and pervasive commitment by a community to a conceptual framework, is mirrored in the allegiances, and imbalance that historically have existed between traditional and performance assessments. The problem exists when paradigms are embraced too fervently: paradigms not only serve to select what is legitimate for study or use, they also specify what is necessarily excluded from acceptability. As a result, many of the findings from the research on the assessment-curriculum relationship have left out the perspectives of those working with educational innovations. It is possible that these perspectives may vary according to a number of characteristics, e.g., the degree to which the participants adopted the assessment innovation on their own, the sense of ownership, and commitment in developing it.

The problem with ignoring participants' experiences in developing and using the innovations is that we learn very little about the context of the innovation, the way it is shaped, and what it looks like after it has been implemented at site level. The research, for the most part, describes rather than qualifies the nature of participants' experiences in developing innovative assessment-curriculum programs.
Chapter V

Methodology

This chapter addresses the methodology used in documenting this study. Issues related to research design, rationale, access, data collection, validity and reliability, analysis, and interpretation are discussed.

1. Research Design

A case study design was used in this inquiry. This study's design was chosen in order to illuminate the emergent meanings, practices and consequences associated with the assessment-curriculum relationships within Language Arts curricula in the third, and sixth grades of two schools. The study's ultimate purpose was to "discover theory from data" through the "general method of comparative analysis" (Glaser & Strauss, 1967, p.1). Guba (1978) depicted the practice of qualitative enquiry as a wave on which the researcher moves from varying degrees of emphasis between a "discovery" and "verification" mode. As the inquiry revealed patterns, I began to focus on verifying and clarifying what seemed to emerge from the data. At this point, data collection and analysis became deductive - moving back and forth between separate elements and combinations of elements, between parts and the whole, in what Kuhn (1982) termed a "sorting-out, putting together process". Later, I attempted to understand the multiple inter-relationships among categories and themes that emerged - an inductive approach to understanding participant activities and outcomes within specific educational contexts.
There are two phases in the design of this study: historical, and explanatory. The historical stage focuses on gathering data regarding 1) the contextual origin, with the various components of meaning which have been gradually attached to the assessment-curriculum relationship in the classrooms studied; and 2) the experiences of teachers, students, and school administrators at each site with regard to current assessment practices, their meanings, and their consequences on classroom practices. Patterns emerging from the data collected were identified and analysed. During the explanatory stage, the plausible explanations for patterns discovered in the description stage were synthesized (coded, and categorized) into meaningful explanations. These explanations were revised according to further data that either confirmed or disconfirmed them. The results are developed into theories and discussed in terms of their implications for other cases, i.e., what one would expect to occur under various conditions if the theories were correct. Finally, conclusions, implications and recommendations were developed based on the analyses and comparisons of the findings from the classes and participants in the two schools studied.

For each research site, the continuous components of the historical and explanatory stages can be displayed as:

i. Historical:

\begin{itemize}
  \item Collect Data \item Analyse Data \item Form Explanations \item Collect More Data \item Analyse Data \item Interpret \item Research Design Evolves \item Interpret
\end{itemize}

ii. Explanatory:

\begin{itemize}
  \item Revise explanations to fit data \item Generate Interpretation
\end{itemize}
Two basic concerns influenced the design of this research approach: the interest in determining insiders' views of the relationship between assessment and curriculum; and the consequences of these relationships on classroom practices in two schools developing the Toronto Benchmark Program. It was critical that between the two schools chosen, there existed a wide range of experiences in developing the Toronto Benchmark Program in order to provide tangible contexts for examining how the curriculum-assessment relationship varies according to stage in program development. Further, the examination of the consequences on classroom practices guided the construction of theory in the explanatory mode of the design.

2. Research Rationale

In an attempt to understand the meanings in developing an innovative assessment program, I focussed on the experiences of participants in two schools. My enquiry respected Metz's (1986) statement that culture "...is not a systematic set of logically interrelated propositions about values, norms and nature of the empirical world", but is also "...a broad, diffuse, and potentially contradictory body of shared understanding about both what is and what ought to be" (p. 54). Further, Erickson (1986) reinforces the aspect of "subculture" as "learned and shared standards for ways of thinking, feeling and acting" (p.117). A range of "meaning systems" operating within and between the shared spaces of schools allows for the possibility of conflict over what and how curriculum is developed. The case study design in this inquiry was consistent with a "cultural" focus on the experiences and meanings of the participants in two schools involved in developing an innovative assessment program. Hammersley and Atkinson (1990) and Woods (1991) comment that only through such inquiry can the meanings that give form and content to
social processes be understood. Therefore, the utilization of grounded theory methods used in this study involves 1) teachers and students participating in the social world of the classroom, and 2) administrators and parents engaging in the products of that participation.

Grounded theory methodology is attractive from the interpretive researcher's stance because of its open-ended relation to the development and testing of theory (Glaser & Strauss, 1967). "Over time the...researcher...has the opportunity to check out his or her understanding of the phenomenon under study" (Hammersley & Atkinson, 1990, p. 24). Here, the phenomena under study are the meanings and consequences involved in developing an innovative assessment program. Therefore, theories formulated from the data collected can be assessed and reformulated in light of new data collected throughout the research process (Wolcott, 1990). This allowed me flexibility in the direction of the research enquiry, and also enabled the participants to engage in the research process as it continued to evolve.

Grounded theory inquiry is particularly suitable for reducing the gaps between researcher and teacher, and between educational theory and practice (Woods, 1991). It concerns itself with what people are, how they act, and why. It attempts to uncover people's beliefs, motivations, theories, values and behaviors, and considers how these might change over time, or from one situation to another. It focuses on an insider view of participants' meanings, experiences and interpretations - thus making available various levels of social meaning in their full richness (Woods, 1991).

Throughout the research process, it encourages the researcher to participate (Hammersley & Atkinson, 1990). For me, not only did this involve acknowledging my particular actions and thoughts while observing
and interviewing participants, but it also assisted me in reflecting upon my actions in those processes. As Wolcott (1990) states:

> What one looks at and writes about depends on the nature of the problem that sends one into the field in the first place: on the personality of the...researcher; on the course of events during fieldwork; on the process of sorting, analyzing, and writing that transforms the fieldwork experience into the completed account; and on expectations for the final account. (p.191)

My participation in this enquiry becomes a valid source of data and impacts on how I construct and interpret data in this case study. As defined by Denny (1978), the case study is "an intensive or complete examination of a facet, an issue, or perhaps the events of a geographic setting over time". In this study, the "bounded system" (Smith, 1978) selected set up the opportunity for viewing assessment-curriculum relationships and their consequences from two vantage points: early and advanced stages of curriculum development. While the two schools differ with regard to location, and timing in their program activities, they share the same goal in that they are both involved in the development of the Toronto Benchmark Program. The case study approach was also conducive to uncovering shared meaning systems that contextualize the actions and interpretations associated with educational processes.

Accordingly, the interpretive (Erickson, 1986) case-study approach has been favorably utilized in this study for four reasons. First, it illuminates the invisibility of everyday classroom life. What is happening with classroom assessments? is clarified. Second, the interpretive approach addresses our need for specific understanding of the assessment-curriculum relationship through documentation of concrete details of practice. Third, the local meanings that the assessment-curriculum relationship provides for the
participants are highlighted, and fourth, it addresses our need for comparative understanding of assessment-curriculum relationships in different social settings.

3. Access

My first step in gaining access to the two sites required for my data collection was to complete the required "Research Application For Toronto Board Of Education". Information that I needed to disclose related to 1) the problem addressed in the study; 2) the purpose of the study; 3) the relevance and potential utility of the study to the educational system; 4) the school facilities and sample required; 5) the dates for data collection; 6) the method and forms for obtaining informed consent; 7) the provisions for preparing subjects; 8) the copies of data collection instruments; 9) the provisions for feedback of study to participants; and 10) a thousand-word description outlining my study's salient features. After my university thesis advisor and department chair critiqued and signed the completed form, I delivered it to the Acting Chief Research Officer at the Toronto Board.

Prior to submitting my request for consideration to the Board's Research Review Committee, the Research Officer unofficially critiqued my proposal in light of criteria for research set out by the Toronto Board. After making some written changes regarding the terminology and research process, the proposal was submitted to the Research Review Committee's early November (1993) meeting. In discussing the proposal, the Committee expressed concerns about the feasibility of my studying the Board's Benchmark Program in light of its very early implementation in most of their schools, i.e., they were concerned that I might not be able to collect relevant data from their schools. At this point, the committee members suggested that other opinions on my
access request be solicited from those school superintendents who might be aware of schools that were in the more advanced stages of development with the Toronto Benchmark Program. Ensuing discussions between the Board's Research Officer and two of the Area Superintendents resulted in permission being granted for my access to two Board schools. However, this access was conditional upon including Toronto Board parents in my study's sample in the two cooperating schools; and upon completing a research report, based upon the results of my study, for the Toronto Board. Originally, the parent sample was not included in this study since it was felt that their presence was somewhat removed from the study's classroom unit of analysis. However, it was felt by all concerned that the inclusion of parents in my study would have more potential for informing and guiding Board, school and community concerns with the development of curriculum and assessment practices in their classrooms.

The principals of Montrose and Arcadia Schools had discussed and solicited participation in my project with their school staffs. I met with them to discuss, in more detail, my research project. I also met the cooperating teachers, and made arrangements for the collection of data. Thus, my request for access began in September and was granted in November of 1993. Because of the busy schedules of my participants just prior to the Christmas season, I began collecting data in early January of 1994.

4. Ethical Considerations

Proper ethical procedures for proceeding with this investigation were taken by filling out the Ontario Institute for Studies in Education "Ethical Reviews: Statement of Intent" form. When data collection began, I informed all participants, either in-person or by telephone, about the nature of the study
and my expectations of them. I also provided regular opportunities for participants to read or to discuss the text attributed to them in order to ensure clarity of thought and accuracy of statement. The participants were asked to sign a copy of the prepared consent forms (Appendices F to I). These forms explained the data collection techniques requested for each participant, the purpose and nature of my study, assurance of anonymity, confidentiality, the freedom to communicate with my advisor on this study, and freedom to withdraw their participation from this study without penalty. In order to preserve confidentiality, the names of the participants in my study - their schools, and their Board, have been changed. Data belonging to the participants have been reported so as to preserve the authenticity of their voices without revealing personal identities. All of the participants in this study from Montrose have been given pseudonyms beginning with "M" while participants from Arcadia have been given pseudonyms beginning with "A".

5. Participants

Four teachers, seven students, seven parents, and three administrators directly associated with the third and sixth grades of Language Arts classes in each research site emerged as the participants in this study. In addition, one staff development officer from the Toronto Board was interviewed. During many of these observations, I acted as a resource to students with their classroom activities, i.e., as a participant-observer.

The principals of both schools saw value in having this study conducted in their schools. Initially, two teachers from each school volunteered to participate in this study. Participants were selected according to their experience in developing and implementing the Toronto Benchmark Program. As much as possible, I engaged participants at each site that were balanced in
gender, teaching experience, and in use with the Toronto Benchmark Program.

In Montrose, all four teachers were female. Muriel was in her twenty-second year of teaching; Margaret in her eighteenth; Michelle in her fourteenth; and Maureen was in her seventh year. All had been teaching in Montrose for at least six and one-half years. Of these four teachers, Muriel and Margaret were instrumental in developing the Benchmark and Portfolio programs at their school. The other two, Michelle and Maureen, while not as involved with its initial development, had integrated the Benchmark Program into their teaching. The principal, Ms. Merton, was in her first year at this school - she replaced the retired principal, Mr. Mathews, who had initiated the Benchmark (Pilot) Program in Montrose.

In Arcadia, three of the teachers (Annette, Adelle, and Arvel) interviewed and observed were female; the fourth (Alex) was male. Of the three females, Annette had 18 years of teaching experience; Adelle was in her seventh year; and Arvel was in her first year of teaching. Alex was in his seventh year of teaching, also at Arcadia. With the exception of Arvel, all were in their third year of development with their Board's Benchmark Program, and all were just beginning to use portfolios with this program. The principal, Ms. Ashton, had been in this school as administrator for the previous three years.

A diverse [ability] range of students were interviewed in each school. They were selected on the basis of having participated in Benchmark activities during the past school year. The main difference was that while some of the students interviewed in Arcadia represented a mixed-cultural group, those in Montrose were more homogeneous, i.e. all were born and raised in Toronto. In all cases, the parents interviewed were those of the students interviewed.
Obvious differences existed between the two groups of parents. Those connected to Montrose resided in the school's immediate and highly affluent neighborhood, were well-educated, and were established native Canadians. Those connected to Arcadia were residentially much more dispersed - some living a great distance from their children's school, in much less affluent neighborhoods, and were mostly new to Canada within the previous five years.

All of the participants were key informants (Bogdan & Biklen, 1982) for the study, i.e., they were individuals who had "a greater experience in the setting, or were especially insightful about what went on" (p. 63). The sample was purposive (Chein, 1982) and criterion-based (Goetz & LeCompte, 1984).

6. Data Collection Methods and Instruments

The data originated from a textual and electronic search of the theoretical issues critical to the assessment-curriculum relationship; and from the analysis and syntheses of the on-site research.

To conduct the literature review, two separate computer searches were conducted. ERIC batch searches covering the years from 1984 to 1997 were run on the major areas of the study. Of particular importance were the following descriptors:

- Assessment
- Educational Accountability
- Educational Improvement
- Educational Innovation
- Language Arts Education
- Program Effectiveness
- Student Learning
- Teacher Instruction

In addition, a hand-search of journals, and texts was made. Sources from these searches (and any other materials) were then used to locate other
pertinent documents. In addition, I used my personal files, and communicated with other researchers who were investigating the assessment-curriculum relationship.

The primary method of on-site data collection was through semi-structured interviews with teachers (Schedule A), students (Schedule B), administrators (Schedule C), and parents (Schedule D). The secondary method of collection was through participant observation (Schedule E). In addition, documentary analyses (policy statements, and researcher field-notes) were used.

A total of 21 (45-minute) interviews were conducted at each of the two school sites to which I was granted access. In addition, the Board's Staff Development Officer was interviewed. Ten class observations were conducted at each site in the grades three and six Language Arts classes. In all, I conducted a total of 43 interviews, and 20 class observations between January and June of 1994.

However, the amount of data collected in this study reflects variations amongst the participants with regard to their use of the Toronto Benchmark Program. Consequently, of the four groups of participants, the teachers and administrators contributed most to the data corpus as they were most directly involved in the development and use of the Benchmark innovation: the students' data were used to clarify how the innovation was operationalized in the classroom settings at each research site; and the data from parents were used least frequently as this group was considered to be tangentially related to the research question of this study. The main sources for the teacher data were Muriel and Margaret at Montrose; and Annette and Adelle at Arcadia. Secondly, because the participants from Montrose were the most experienced with the development and use of the Benchmark innovation, they constituted a broader context in this study, i.e., they were able to provide more data that
were focussed on the antecedents, transactions, and outcomes associated with the innovation.

**Data Collection Methods.** This inquiry used five methods of data collection: i.) interview; ii.) observation; iii.) documentary analysis; iv.) researcher field notes; and v) member checks. A description and detailed rationale for each of the four methods follows.

i. **Interview.** The interview was an appropriate technique for this research as "The ability to tap into the experience of others in their own natural language, while utilizing their value and belief frameworks, is virtually impossible without face-to-face and verbal interaction with them" (Guba and Lincoln, 1981, p. 155). Through interviews, researchers can clarify what they thought happened, achieve a fuller understanding of the incident in question, and also take into account the lived experiences of the participants. The purpose of interviewing, then, is to allow us to enter into the other person's perspective.

In this study, the strengths of the interview as a data collection method were significant (Merriam, 1988; Gay, 1981; Guba and Lincoln, 1981). Arrangements were scheduled to the mutual satisfaction between myself and the participants several days in advance of each meeting. The interviews were flexible enough to allow the interviewer to adapt to the temporal, and professional constraints of each participant. They provided a high response rate, and offered the possibility of eliciting in-depth data from the participant involved in the interview. In establishing rapport, I was able to obtain data that would often not be related directly to the interview schedule. People were more apt to disclose themselves, their thoughts, feelings, and values than they would have in a less personal situation. The interviews resulted in
straightforward and accurate responses since I was able to explain and clarify both the purposes for the research, and the questions being asked. Incomplete, divergent, or unclear data was followed up through a request for elaboration or clarification. In turn, the interviews served as a validity check for data provided through observation, and documentary analysis. While all interviews were tape-recorded as they occurred, notes were made of the interviews within twenty minutes after they ended.

However, particular weaknesses with the interview method need to be addressed. Inferences about validity may be made on the basis of questionable (face) measurements that may or may not measure what they claim to measure. Bias, a tendency to make errors in a particular direction, may also threaten the validity of a study. Sources of bias may be related to the characteristics of the interviewer, those of the participant, and/or the content and wording of the questions.

In this study, steps were taken to reduce the elements of error due to incorrect inferences. There was careful formulation of questions so that their meanings were made as clear as possible. Validity of the interview data was checked through pilot testing with students, teachers, parents and administrators; assuring participant confidentiality; full recording of the interviews; and re-checking and cross-checking of divergent data. The reliability of interview data was increased by cross-checking data derived from observation, and documentary analysis. Data received from one person were presented to another in order to establish cross-participant agreement on its accuracy. In addition, requests for participant acknowledgement, and analyses of data helped to promote reliability in this study.

ii. Observation. The second method used in this study was participant observation. I observed, and participated in the setting so that I was part of
the social dynamics taking place (McCall & Simmons, 1969). In addition, I acted as a resource teacher with students in their grades 3 and 6 Language Arts classes by assisting the students with their Language Arts classes.

The advantages associated with using participant observation in research build on direct experience (Douglas, 1976). Inquirers are placed in the position of having to ask others about practices and events that they themselves did not witness - we are usually most convinced by our own direct experience with a situation, thus providing a 'test of truth' for that situation. It allowed me to record behavior and events as they occurred. The absence of a time lag between observation and recording contributed to the accuracy and validity of the data observed. Observational techniques made it possible to build on both propositional and tacit knowledge, enabling me to interpret the broadest range of inputs. This aspect of observation maximized discovery and description. Further, when distance, memory lapses, or emotional reactions may cause significant alterations in the data that I was seeking, observing events for myself became an added precaution against bias. In addition, participant observation maximized my ability to grasp motives, beliefs, concerns, interests, unconscious behaviors and customs. It allowed me to see the world as my participants saw it, to live in their time frames, to capture the phenomena in and on its own terms, and to grasp the culture in its own natural, ongoing environment. I was offered emotional reactions of the group introspectively - permitting me to use myself as a data source. As well, it allowed me to build on tacit knowledge, both my own and that of the members of the group.

Observational techniques may lead to reactivity in the setting or on the part of the members (House, 1978). The method leans heavily on personal interpretation. As a result, direct experience with or involvement in the
situation might lead one to experience a setting in a biased manner, and lead to self-deception on the part of the investigator. This close involvement may cause the investigator to take meanings for granted, and therefore make it impossible for him/her to observe or report on them.

Precautions were taken to counter the arguments against participant observation. I experienced prolonged engagement on-site. Repetition of observations was used to give a high degree of acquaintance with the salient activities, and to provide an opportunity for those being observed to become acclimatized to my presence in the classrooms. All of the participants were assured of anonymity and confidentiality. Observational data were triangulated with interview and documentary data. As well, I checked data and interpretations with various members of the group from whom data were received throughout the study.

iii. Documentary Analysis. Documents and records that related directly (primary), and indirectly (secondary) to this inquiry were used. For example, those that related directly included *Benchmarks: Standards of Student Achievement*, (Grades 3 and 6 Print and Video Benchmarks, Toronto Board Of Education); *Using The Benchmarks: A Handbook For School Staffs* (Toronto Board Of Education); and *Benchmarks: The Development of a New Approach to Student Evaluation*, (Sylvia Larter, Toronto Board of Education, 1991). Documents indirectly related to this enquiry were *Observing Children* (Toronto Board of Education, 1980); *Language Across the Curriculum Guidelines for Schools*, (National Association for the Teaching of English, 1976); *Conversations With James Britton*, (Toronto Board of Education, 1974); *Teaching English And The Language Arts: A Policy Paper* (Toronto Board of Education, Language Study Centre, 1982); and two staff and parent council meetings at each school site.
Documents used provided a stable, rich and rewarding resource. They were a base from which I could work and thus offer stability to this study. Many of the documents embodied a policy base upon which to ground data. They represented a 'natural' source of information, i.e., they showed what happened in the context, and provided a variety of evidences and perceptions about the environment. They also lent contextual richness and support in grounding the enquiry in the milieu of the participants, and events under investigation.

The direct documentary sources presented subjective views of events in that they recorded individual impressions of the events associated with the inquiry. Otherwise, it may have been difficult to obtain a fully developed view of the event under study. Indirect sources required inference on my part. Since the documentary source was being used for something other than what was originally intended, the inadvertent sources often tended to contain information irrelevant to this study. The reliability of data gathered through documentary analysis was assessed by examining them for internal and external consistency, i.e., establishing the extent to which the data were consistent within and across documents, and across data collected from other sources about the same event. I conducted checks on data collected through documentary analysis by comparing them to data collected through interviews and observations.

iv. Researcher Field Notes. I wrote my own account of the data-collection sessions. To fit into the research setting, and to record data that were accurate and comprehensive, I dictated these field notes into a tape-recorder in an area separated from the participants' environments. The notes were descriptive, concrete and detailed in relation to the context, events, and unexpected occurrences during data collection. I recorded my reflections soon after each
data-collection session, and placed them in my notes. They were subsequently transcribed, and analyzed in the order received. Often in direct quotations, these notes contained what people said in relation to some practice. Because part of the purpose of participating in both settings was for me to experience the characteristics of each, the notes often contained my own reactions to the characteristics, and to the reflections about the personal meanings and significance of what occurred. These notes complemented the data corpus - they guided and informed me as to the evolving patterns in the inquiry across times, participants, and sites. Finally, the field notes included my insights, interpretations, and early analyses of what was happening in each setting.

Thus, steps were taken to counter threats against reliability and validity with the field notes. For example, early transcriptions and member checks matched the accuracy, clarity and relevance of the notes to the intent of this study.

v. **Member Checks.** Feedback data were collected from the participants at both school sites by inviting them to correct errors of fact, and to supply alternative explanations or modifications to the theory developed by me. This process helped to assure validation of my findings and evolving theory development (Miles, 1980).

7. **Interpretation**

The interpretive framework constructed by Erickson (1986) was used as the structure within which the study examined the voices of the participants associated with the development of the Toronto Benchmark Program. Erickson defined the criterion of validity as addressing the *immediate and local meanings of actions* from the participants' points of view. Intrinsic to his model is the role that interpretive research plays in furthering
understanding of concerns central and substantive to classroom teaching: a) the nature of classrooms as socially organized environments for learning; b) the nature of teaching as an aspect of the reflexive learning environment; and c) the nature of the meaning-perspectives of teacher and learner as part of the educational process. Erickson's initial step was to develop a set of five questions (Table 5).

Table 5
First Framework for Interpretive Research

1. What is specifically happening in social action that takes place in this particular setting?

2. What do these actions mean to the actors involved in them, at the moment the actions took place?

3. How are the happenings organized in patterns of social organization and learned cultural principles for the conduct of everyday life i.e., how do people interact with others to create meaningful actions?

4. What is happening in this setting as a whole, i.e., the school building?

5. How does the organization of one setting compare with other ways of organizing social life in other places and at other times?

The study reported here is, to the researcher's knowledge, the first to attempt to apply the concerns hypothesized by Erickson to the realities of the participants associated with the development of the Toronto Benchmark Program. For purposes of meaningful dialogue, I found it necessary to develop a modified version of the framework appropriate to fulfilling the purpose of this study. Each of the five questions of Erickson's (1986) model were field-tested on administrators, parents, teachers, and students. Based on
concerns for a) conceptual clarity; b) participant understanding of research issues and interview questions; c) relevance to participant activities; and d) public controversy, this framework was modified to that seen in Table 6.

Table 6
Second Framework for Interpretive Research

I. Meanings and Purposes of Benchmark Assessment:
1. a) What is your interpretation of the terms:
   i) classroom assessment,
   ii) Benchmark Program, and
   iii) curriculum, e.g., your classroom activities, the subject to be taught
b) How do you think the interests of students, teachers, and parents are served by Benchmark assessments?

II. Issues with Developing the Benchmarks Program:
2. a) How have you helped develop Benchmark assessments for your classroom?
   b) For you, what have been the major issues in developing Benchmark assessment?
   c) In terms of every-day use and educational support, how would you compare your experience between your current Benchmark and any previous assessment?

III. Classroom Activities and Assessment with Benchmarks:
3. a) How would you describe the Benchmark assessments you use in class? (e.g., do they relate to the teaching of skills, knowledge, attitudes, participation?)
   b) What information do Benchmark assessments give you regarding your students, your teaching, and your course? (e.g., how you're meeting the course objectives?)
   c) In what ways do you use this information? (e.g., change the way you do your teaching?)
   d) Are you personally satisfied with the usefulness of this information?

IV. Interaction between the Curriculum and the Benchmark Assessment:
4. a) What influence do you think your curriculum (e.g., language) has on your Benchmark assessment? (e.g., have they become more "discrete"?)
   b) In turn, are there ways in which your classroom assessments have influenced the courses that you teach? (e.g., the content, objectives, instruction)
c) How does the influence on your school curriculum and assessment exerted by your Benchmark assessment compare to that experienced with non-Benchmark assessments?

d) What kind of relationship would you like to see between the course you teach, and the way you need to assess your students' performance? For what reasons?

V. Educational Consequences of the Benchmark Program:

5. a) Can you name any educational benefits and/or weaknesses derived from your Benchmark assessment on:
   i. effects within your classroom? (e.g., student-teacher relationships, motivation?)
   ii. effects within/on your school? (e.g., communication with other teachers, parent reporting, professional development?)

b) How would you say these consequences compare to those you may have experienced before you used the Benchmark assessments?

From a subsequent check to ensure that the concerns of those pilot-tested and those of the researcher concerning workability were met, the framework evolved into that seen in Table 7.

Table 7
Third Framework for Interpretive Research

1. (Context). Can you describe the environment within which you developed the Toronto Benchmark Program?

2. (Meanings). What does "Toronto Benchmark assessment" mean to you? What do you think the purpose of this assessment is?

3. (Issues). What do you think were the issues involved in developing the Toronto Benchmark assessment?

4. (Relationships). How would you describe the relationship between school curricula and Toronto Benchmark assessment? (i.e., how do they work together?)
5. (Consequences). From your experience, what were the educational consequences of this relationship:
i. within the classroom, and
ii. for the school?

I added the component of "context" to Erickson's framework so that participants' actions and experiences could be linked to their specific educational environments, and to the issues in the literature found critical to assessment-curriculum relationships and innovations.

Underlying the framework for interpreting participants' perspectives in this study was the naturalistic or interpretive paradigm (Cohen & Manion, 1980; Guba & Lincoln, 1982; Smith, 1978). This paradigm assumes that knowledge needed to understand human behavior can be acquired only through inquiry into the dynamic nature of social interaction. There are multiple, intangible realities, and these can be studied only in a holistic manner, since to fragment the whole would be to cause significant alteration of the realities. The aim of this approach was to search out modes of explanation based on data found in the situations themselves. The naturalistic paradigm was judged most appropriate for this study, since the purpose was investigation into the social and behavioral interactions involved in adopting and developing an innovative assessment program. The data were then used to describe, analyze, and conceptualize that phenomenon.

8. Data Analysis

This study's data analysis was driven by the focus on its research question. Data analysis began almost simultaneously with its collection, and in much more detail once the transcripts from the observations, interviews and
documentary analyses were completed. Within the theoretical framework of the assessment-curriculum relationship, data were initially organized according to three basic, but broad sections: 1) time (the historical and current context of the assessment-curriculum relationships); 2) personnel (teachers', students', parents' and school administrators' practices, meanings and views of assessment and curriculum); and 3) site (developed and early-stage implementation within Toronto Benchmark schools). The data were coded, clustered, and classified according to the numerous themes that were interpreted in this study. In using the grounded theory method (Glaser & Strauss, 1967) of data analysis, I worked through three stages of data analysis. Stage I involved words that described the action in the settings. Exact words were rewritten on the widened right-hand margins of the transcribed data texts. For example, such substantive codes as "preparing", "integrating", "adapting", "focussing", "referencing", "conferencing" based on and extracted from the data were not only meaningful, but also helpful in preventing me from imposing my own preconceived impressions on the text. On the left-hand margins of the data texts, variables, e.g., "context", "consequences", that corresponded to Erickson's (1986) framework were noted. In addition, codes not found in Erickson's framework, but which were located in text, were recorded under the heading of "other". Stage II resulted from condensing Stage I codes. As more data was collected, the themes became increasingly obvious. Decisions about categories were made by asking questions of the data. For example, I asked what was meant with the statement "The role of the teacher hasn't switched properly"; compared the statement with others that were similar, and then asked myself "What category would similar incidents fall into?" Finally, the researcher compared each emerging category with others to ensure that they were mutually exclusive and covered all behavioral
variations within each category. During Stage III, I conceptualized the theoretical constructs among the codes, 'weaving the fractured data back together again' (Glaser, 1978, p. 116).

The naturalistic design of this study provided opportunities for comparative analyses on several levels. The responses (i.e., congruencies and discrepancies in perceptions and experiences) were compared within and across four classes of participants (teachers, students, parents and administrators) from the i) developed; and the ii) early-stage implementation Benchmark school sites. The data were analyzed further to provide possible explanations for any exceptional events observed and reported. Glaser and Strauss (1967), and Strauss and Corbin (1990) were referenced as the sources for the qualitative analyses of data in this study.
Context of the Montrose and Arcadia Research Sites

Researchers (Goodson, 1992; Patton, 1985) comment that the background of participants in a case study helps not only to shape their views of teaching, but also to influence how they conduct their classroom practices. Patton (1985) comments:

...what it means for participants to be in that setting, what their lives are like, what's going on for them, what their meanings are, what the world looks like in that particular setting - and in the analysis to be able to communicate that faithfully to others who are interested in that setting. (p. 1)

This chapter traces the contextual characteristics of the Montrose and Arcadia research sites. the Toronto Board's shift to performance assessment, and the characteristics of the Benchmark innovation.

1. Montrose and Arcadia Research Sites. Montrose is a 450 student junior school situated within a very affluent section of the City. At the time of this study, it was in the second of a two year pilot project with the Toronto Benchmarks Portfolio Program. Along with the administrators, many of the 22 teachers at this school have been on the ground floor in conceptualizing and developing the Toronto Benchmark and Portfolio Programs currently in use across the Toronto Board. Teachers from Montrose now provide professional development services to staff in neighboring schools in order to help them
implement the Program. Students in Montrose have daily access to its library's rich supply of print, audio, and visual learning resources.

Muriel from Montrose recently gave a Board inservice on motivating students. Some of the teachers in attendance were from Arcadia. After introducing the topic of the inservice, one of the teachers from Montrose commented:

Oh, its okay, we don't have motivation or attendance problems with our students. They're always at school, always on time, and are very interested in their work. They're a pleasure to work with.

Annette from Arcadia adds to the description of teachers at Montrose in terms of their work with the Benchmark innovation:

And the teachers...you won't find another school that for the Benchmarks has so much support, so much collaboration and team work, so focussed! They're just so together about it.

Although administrator Mr. Mathews of Montrose retired from service during the time the Benchmark Program was being put into place, his successor (Ms. Merton) was equally committed to this program. Therefore, data from both of Montrose's administrators are included in this chapter.

Alternatively, the second research site in this study (Arcadia) is a junior/senior public school in inner-city Toronto. It houses 485 students from kindergarten to grade eight. Together these students make up a very diverse multi-cultural population. Twenty-seven teachers in the school provide a core program of Language Arts, Mathematics, History, Geography, Guidance and Visual Arts. Special Education services are provided through learning disabilities and Learning Centre programs. ESL support is available for "New
Canadian" students. For the 1994-95 school year, the implementation of the Benchmarks Program was one of Arcadia's main focuses. It was intended that Benchmarks in language were to be incorporated into the curriculum at all levels, and that the use of the Benchmarks Language Arts portfolios in reporting to parents would continue to be refined and extended. Arcadia has a reputation for academic excellence and respect for student diversity. Adelle described the site as an inner-city school:

We have almost every ethnic group in the country here at all economic levels. Some of our kids spend an hour on the bus getting to and from school. Many of them are latch-key children who have to hang around their door for their older brothers or sisters to get home before they can get in. The school population is very transient...most students are here for only two years, some for less than six weeks - and they're gone!

In effect, Arcadia is not a "neighborhood" school in terms of its immediate catchment area. The residential areas of many of these students differed from the Arcadia neighborhood in terms of its social and economic characteristics. Nevertheless, the school presented a mission statement common to all of its students that focussed on fostering academic development, critical thinking, and adaptation to change. It also sought to provide its students with a learning environment focussing on trust and diversity, and which promotes students' self esteem and confidence.

The themes coded in describing the context of both research sites help to create the starting-point of each research site in regard to assessment and teaching. The themes common to both sites centre on 1). introduction to the innovation; 2). administrative experience and educational insights, 3). professional concerns, 4). documents used, and 5). approach to instruction.
I. Introduction to the Innovation. Montrose administrator Mr. Mathews explains how he and his staff had been working on merging instruction with evaluation for three years when the Toronto Benchmark Program was introduced to them. He explained, "We latched on to it right away. The Benchmark Program is a very compound...a very complex kind of thing...it is both program and evaluation". Further, he links his staff's developmental approach to learning to their "natural fit" with the Benchmarks:

No one said "You have to change programs". Instead the staff saw it as "What we're going to do is watch, and begin to question our own practices". Once that happens, we're away to the races. We were well into that. When the Benchmarks came in, it was just a natural fit with what we were already doing.

The "natural fit" between Montrose's adoption of developmental learning, and whole language instruction paved the way for its teachers to adopt the Toronto Benchmarks assessment program.

Comments by teachers in the Arcadia site refer to their board office's presentation of the program. Annette gives her reaction to the Benchmarks being presented as "the" instructional program:

The board just hasn't done a good job of introducing this program to us. They've unloaded this package on us and told us "This is what we're going to use from now on...this is really good...Here, mark holistically! This is the way to do it...this is the program". They just threw it at us.
Adelle discussed how she thinks the program might have been introduced in order to receive a higher degree of teacher acceptance:

They [the board] could have said to us: "We have a new program we think is really good for your students and you. You're professionals, and we would like you to be involved with us in the development of the Benchmarks curriculum". If they introduced it to us that way, we might have been able to [accept it more]...

Further, Adelle explained three important issues she thinks needed to be discussed at the time the program was being introduced to teachers:

We needed to discuss where we began [with curriculum and assessment]; Where do we want to get to with it?; and How do we get there? We [the school staff] just never had the opportunity to discuss and reflect on these issues.

ii. Administrative Experience. Mr. Mathews who began to develop the Benchmark innovation at Montrose, taught in both junior and senior schools in the 1960's. After being a math teacher, he went to a school as vice-principal which was associated with the Language Studies Centre. He explained:

I went in to the Language Studies Centre as a consultant, and that's where I became involved with, began to develop into developmental programming...because the consultants were to become links with Vygotsky, Briton, Piaget ...in some depth.

Mr. Mathews' first principalship was in a new K to 3 school. He explained his exciting introduction to student learning as "It just threw me into a whole new world...I mean it was just primary kids...I had to get involved in how young children learn. It was really an exciting time!" He explained that the neighboring grade 4, 5 and 6 school was very traditional. "There was no communication whatsoever...the two staffs had never met each other in all
their years there!" After the Board combined these two schools into one, Mr. Mathews was appointed as its first principal. He said that in his first year:

...we switched half the junior staff, and half the primary...and mixed them so they were side by side. It was just the physical...and people started talking to one another! I say that as a background because the Benchmark Program was really an attempt to evaluate the developmental process of education.

After that, Mr. Mathews explained that when he went into other schools, he began to recognize that there were isolating divisions in each. He cites an example provided by a huge open-area school he visited:

There were three floors. The first floor was the kindergarten and the administration, the second floor was all primary, and the third floor was all junior. They had separate recesses...that had separate staffrooms and staff meetings...it was like the kids graduated into a different room!

The divisions made an impact on Mr. Mathews. When he was the principal of the first Inner-City Project school, he hired "occasional" staff so that his regular teachers could be brought together for ongoing inservices in developmental learning. He explained:

They [speakers] never talked about specific strategies, and I was able to see, first-hand, the impact of that kind of thing. It convinced me: don't start with the program, start with how you understand how kids learn, and then we'll talk about your practices.

Alternatively, Ms. Ashton had been at Arcadia since the Benchmark innovation had first been introduced there in 1991. Previous to that she taught students in early years and special education classes. As well, this
administrator served on the Board's Benchmark Development Committee for two years, and took part in the five-day Leadership Training Workshop designed to help administrators introduce the Benchmark Program to their school staff. She recollected:

We did that for a whole week...trying to think about the whole process of teams, and team-building and leadership. So we got exposed both to the theory of the Benchmarks and how they worked, as well as how to involve staff in the change process.

iii. Professional Concerns. Both of the Montrose administrators had concerns fundamental to their commitment in developing the Toronto Benchmark Program. When introduced to the Benchmark program, Mr. Mathews commented:

I was always bothered by...the kind of looseness of the evaluative part of a developmental program. The Benchmark Program offered new possibilities...you were able to look at it, see it, and put your hands on it. and that was really exciting for me.

Further, he wanted to pursue a collegial. and context-specific approach to change:

We've got to get away from this notion of "top-down" stuff - and start saying "What is it that's making the really good primary programs so good?" and "How do we continue that...how do we build on it rather than go off in a different direction?"

Montrose administrator Ms. Merton also focussed on the identification and use of valuable programs. She says "I think its important for us to find out what it is that makes assessment programs valuable...what makes them work". More specifically, she viewed the Benchmarks as being very supportive to teachers in the areas of assessment, and curriculum as she believed they
provided teachers with examples of good curriculum ideas for classroom use.
Ms. Merton also believed the Benchmarks address the issue of measurement for instruction, and of standards for student performance:

The Benchmarks give us a standard that we can measure by and become more consistent throughout the system...Although schools have a fair amount of local autonomy, there's still the need for accountability and some sort of system standard. So when the Benchmarks were developed, they got a sample of what grade three's were capable of. So it certainly gives us a standard from which to assess our children city-wide.

Ms. Merton commented that the levels of performance associated with the Toronto Benchmarks Program provided teachers, students, and parents with a base for growth:

I believe in knowing the child's development and progress in much detail...so instead of the generalization "so and so is doing well in language", we can say "so and so can ..." and we know what the next step in the learning continuum is...so we can move students in smaller increments and figure out clearly where we're going. It helps teachers plan their instruction, and communicate to parents about their childrens' growth...

At Arcadia, Ms. Ashton believed strongly in the need to locate the actual performances of students if we are to assess them credibly:

We really have a lot of work left with these assessment programs. But, I don't think there are any quick fixes...I think our assessments need to be grounded on our students' actual performances, and there's no getting around it...we need to find out what these actual performances are if our assessments are to be credible...if our schools are to be accountable.
Ms. Ashton also believed that the processes and activities in the Benchmark Program are those that she would like to see taking place in her teachers' classrooms:

The Benchmarks are not separate from the assessment and they're not a special thing...not anything extra. So those things that the Benchmarks look at are those activities in language that the teacher should be doing in their classrooms all the time. In terms of assessment that takes in teacher observation, formal and informal testing...teachers can take a piece of work and do an evaluation on students.

She saw the Benchmarks as flexible, and stated broadly enough for application across many grade levels. The Language Arts coordinator for Arcadia added:

My interpretation of it is that's it a series of examples of good activities that one can use in one's program that can be adapted and can be used with other materials...but what's different about them just being a list of good activities is that they have criteria attached to them for evaluating kids in using those activities.

When asked what she saw as the major purpose of the Benchmarks, Ms. Ashton commented:

Well, I see it in terms of both improvement to instruction, and also in terms of evaluating a child's performance on those activities, and instruction that occurred in the classroom, because in the past when we used standardized tests, we may not have been testing what we were teaching.
When asked if the alignment between instruction and testing related to school accountability, Ms. Ashton identified the linkage between the two:

Definitely! And I think we need to be accountable, and parents are right in saying "I want to know where my child is [academically]."

iv. Documents Used. Administrator Mathews from Montrose explained that when the report on Measurement, Evaluation and Reporting to Parents was drafted in 1979, concern was expressed by staff at Montrose regarding the lack of specific information available for teachers' classroom use on observing children:

We had objected to a committee that had been set up by the Board to look at evaluation. All they did was look at the different kinds of tests that you could give kids, and we said "You're leaving out a whole thing of systematic evaluation of kids through which their work can be judged". So we said "We want systematic observation included". They said "Fine, you go get the material on systematic observation and we'll include it".

However, even with the help of the librarians, little material on systematic observation could be found. Mr. Mathews concluded it did exist, but only "...in the heads of the individual teachers!" This realization motivated board principals and researchers in 1980 to publish Observing Children. The document offered assistance to teachers those procedures of student growth not readily addressed by standardized tests and paper-and-pencil exercises.

The importance of Observing Children is frequently referenced by administrators and teachers alike. Montrose's administrator explains that "...it was the first attempt by a group of people to look at assessment as other than testing". He also believed that it set the conceptual groundwork for
identifying the expectations for student performance in Language Arts. For example, the content material and performance levels in the Toronto Benchmark Program is based on the observation of student activities across the Toronto Board. As pointed out by the Board’s Language Arts consultant for Montrose:

We worked on developing Language Arts tasks, and then we actually put teams of teachers together in the summer. They watched, read and and critiqued what the youngsters were doing in writing...and we had a lot of examples, and from that we drew out the criteria for assessment from what the kids were doing.

As well, administrators and staff from Montrose used the document Language Across the Curriculum by James Britton for conceptualizations regarding establishing language as the underlying basis for instruction.

At Arcadia, Adelle explains how she developed a tracking book which she called At a Glance. In her personally developed book, she recorded the names of students, of assignments completed, dates and marks, and extra help needed by her students.

I found this tracking system really helped...just to have things down on paper gave me information on which students needed a more "hands-on" approach.

Alex mentioned Task Force Five (an article depicting the teacher role in North America changing from that of instructor to care-giver). He commented that in Arcadia:

I have to be concerned with my student's social skills...how they interact, their manners, their attitudes to work...The knowledge part is pretty secondary to them.
v. Approach to Student Performance. This subsection of data speaks to the underlying basis upon which the teachers in each research site developed their classroom instructional strategies. The teachers' approaches to student performance were seen to be one of the factors critical in adopting and implementing the Benchmark innovation. In turn, the approaches to student performance related closely to the consequences for the instructional and assessment practices in each research site.

Muriel from Montrose looked at her students' performance based on the principles of developmental learning and whole language instruction. However, she referred to the issue of standards in relation to her students' development:

There are standards out there...what's important is the development of growth in each individual...all people do grow and learn at different rates...we can have standards at the end of different divisions, but that what really matters is the growth and development...

Margaret's instruction was tied to developing a child's abilities starting from where they're at rather than just comparing students' performances to each other by way of numbers:

When children come to me at the beginning of grade one, no two of them are at the same place in development. So it is unrealistic for me to expect them all to end up at the same place at the end of grade one. Some of them are beyond what normal standards I would have had for them at the beginning of the year. So those kids shouldn't be expected to sit there and do those tasks over and over again simply because its grade three material. On the other hand, some kids aren't quite ready for grade three material, so I have to bring them up to speed as best I can while I keep challenging those at the other end of the scale.
Michelle explained that she liked to give a pretest to establish the level of knowledge her students might have with regard to a particular task. Her main purpose and focus, was in developing students' individual growth. She tries to balance her instructional program in order to meet a wide range of abilities and interests amongst the students in her classroom. In contrast, Maureen focused on her students' depth of understanding in the "real world":

I'm not saying that in the real world people sit and read a book and write a letter in role, but they read books, they have points of view of characters...they interpret characters...

At Arcadia, when asked about her approach to instruction, Annette replied:

My approach is basically whole language. Its child-centred, I base my teaching on the personal experiences of my students (for example, if they've travelled, lived or come from the Caribbean, can speak more than one language)...I have the child take personal ownership for his/her classroom activities (for example, its up to each student to try and figure out and put the appropriate effort into making a good cover the student makes for the novel assignment...I take into account the needs of the child how well the student can understand, read or write English), and I ask and use suggestions from the children themselves...some of them might have gone on an interesting trip and would like to write about...do a project on that place.

Adelle explained that she used themes in her class to teach Language Arts, e.g., Olympics, Halloween, Christmas:

I have students write, publish stories, share ideas...So, I use the whole language approach, and structure my lessons and activities according to the level the students are at.

Alex stated that what really mattered to him was in creating a successful lesson "...choosing and providing students with the appropriate reading and resource material". For him, the appropriate content was closely related to engaging
student interest in the topic and motivating them to apply themselves to the tasks-at-hand. Alex felt this was no extra burden on him as he routinely changed his curriculum to suit the needs and interests of his students.

Arvel taught a group of special education students. She looked at the level her students were at, and taught from that point on. She believed the parents of these students were not able to offer appropriate academic support for them. Her main goal was to develop in her students the academic and social skills that would allow them to be integrated into regular classes.

While the two research sites in this study were in the process of developing the Toronto Benchmark Program, their contexts differed markedly. The differences critical to the context of each rest on the administrative approach taken to include staff in the development of the program, the support provided for the development, and the path that each site chose to construct for itself so that it could merge assessment with curriculum.

2. Shift to Performance Assessment. In the 1960's, a growing number of teachers, parents, administrators and students in the Toronto Board considered the more traditional approaches to student assessment to be inadequate. Since that time, the Toronto Board of Education chose not to evaluate individual students against system-wide standards. A number of factors emerged to help make this decision.

First, traditional assessments (such as standardized tests, end-of-course tests, and centrally-developed and scored provincial examinations for graduating certification) were considered inadequate because they were seen as emphasizing the "products" and not the "processes" of learning. At the same time, provincial curriculum guidelines emphasized all aspects of thinking, problem-solving, and process skills, e.g., viewing, listening, role-playing,
speaking, creation of media, estimating, measuring, discussing and debating. Traditional tests were seen to give an incomplete picture of what students "can do". To reinforce the point, the Ontario Ministry of Education in *Issues and Directions* (1980) commented:

Recognizing the diversity of individual abilities and interests, the Ministry views the learner as an active participant in education who gains satisfaction from the dynamics of learning. The concept of the learner as a mere processor of information has been replaced by the image of the self-motivated, self-directed problem-solver, aware of both the processes and uses of learning...The image also reveals a methodical thinker who is capable of inquiry, analysis, synthesis, and evaluation. (pp. 2-3)

Quoting from *Common Sense and Testing in English* (National Council of Teachers of English, 1975), the Language Study Centre of the Toronto Board of Education wrote that the:

...ability to deal creatively with ideas and to express them in an aesthetically pleasing manner is more difficult to measure within the closed format of objective tests. We cannot use objective tests to tell the extent to which students use the resources of the English language to express themselves effectively. (p. 3)

In their 1976 position paper on standardized testing, the Toronto Board wrote:

The Language Study Centre is opposed to the use of the commonly accepted standardized tests as the sole instrument of measurement of a child's ability in language...Teacher observation, informal testing and cumulative samples of pupils' work are necessary to predict ability, level of functioning and potential in language. (p.1)
Similar points of view were found in the educational literature. For example, Valencia and Pearson (1987) articulated their position:

As long as reading research and instructional innovations are based upon one view of the reading process while reading assessment instruments are based upon a contradictory point of view, we will nurture tension and confusion among those charged with the dual responsibility of instructional improvement and monitoring student achievement. (p. 727)

Second, traditional assessment were seen to replace the professional judgments of teachers, even though teachers have the most intimate knowledge of the students. The Toronto Board felt that there was no replacement for a good teacher in the evaluation of students, adjusting programs to meet the needs of each individual child, and reporting to parents. The issue was one of the control of teaching and evaluating, i.e., the improper handling of standardized tests (externally developed, scored and interpreted) may serve to negate these important teacher roles. Madaus (1985) feels that in most instances, ..."the public and its policy-makers mistrust the teacher's judgment. They demand the surrogate mechanical, scientific credentials of standardized test results for the decisions they imagine teachers to be unable or unwilling to make" (p. 615). In short, traditional assessments were considered inadequate because their emphasis on the psychometrics of interpretation were seen by the Toronto Board to attract more attention than did the students who took the tests.

A second factor which helped to sustain the shift in student evaluation was the cultural makeup of Toronto. The Toronto Board has been the chosen destination for immigrant students from many parts of the world. Surveys (e.g., Wright and Tsuji, 1984) found that over 50% of the students in Toronto
schools came from homes where English was not the first language, and that there were over fifty other "first" languages in many school areas. Given this multilingual student population, concerns over test bias arose with the use of standardized testing. For example, many standardized tests were composed of multiple-choice questions that were believed to create a bias related to language and the ability to read and understand them. Cummins (1984) maintains that standardized tests support deficit and transmission models of instruction, create artificial categories, and put minority students at a disadvantage:

...the persistence of abuses in psychological testing reflects a compartmentalization of concern on the part of policy-makers, in that assessment issues are considered separately from pedagogical issues...In this regard, ... current assessment and pedagogical practices conspire to conceal the fact that many of the so-called "learning disabilities" identified among minority students are pedagogically-induced. (pp. 266-267)

As well, many cultures in Toronto expressed concern that standardized tests were sometimes used to limit educational access for students. This concern is well documented in a Toronto Board of Education report *Education of Black Students in Toronto Schools* (May, 1988):

Black parents feel that the statistic of the Board's research that 28.5 percent of Black students were attending schools offering basic level programs gives credence to their perception that Black students are being directed to such schools. Instead of offering Black youngsters an opportunity to "catch-up", streaming they say, often relegates Black students to lower levels of education with lowered expectations. (pp. 26-27)
The parents felt that when their Black children were streamed, they continued to choose the less challenging learning levels in the educational system. The report contained sixteen recommendations related to the topic of streaming, several of which concerned evaluation. For example, here are two of them:

- That the principal in each school review the practices and procedures for assessing and evaluating student achievement to ensure a) the appropriate placement of each student, and b) the remediation and enhancement of each student's "basic skills" as is appropriate.

- That each principal review the procedures for reporting student achievement in his/her school to ensure that there is a plan in place to keep parents and guardians well informed of each student's progress. (p. 30)

Through its report *Evaluation, System Standards and Reporting to Parents* in June of 1987, the Toronto Board initiated a system-wide plan to monitor student achievement and assessment (Larter, 1991). By the fall of 1987, the development of the Benchmark Program was launched. A team of 15 coordinators, consultants, principals, vice-principals, and teachers was formed to develop evaluation methods and materials for classroom use with Benchmark Language Arts programs in the Toronto Board. It was developed to address the need for defining curriculum standards, and for determining student achievement.

For purposes of credibility, it was important that the Toronto Benchmark Program be the "little picture" in this enquiry - the vehicle for locating the consequences of an innovative assessment program. Not only was this program the prototype for Benchmark programs in other educational jurisdictions, but the program was the only one developed from the
observations of local students at work, and from the statistically reliable, and random sampling of students for each Benchmark task. For example, in the grade 3 Language Benchmark (L3-1), a random sample of 360 students were asked to read a poem interactively, and respond orally to questions based on the poem. Their evaluated performance results showed that of this sample, 9% were in Level V (good to excellent), 31% were in Level 4 (good to average); 30% were in Level III (fair); 17% were in Level II (poor); and 13% were in Level I (zero to very limited response). The sample size assured that the Benchmark performances of the sample groups were accurate to within +/- 5% of the population out of which they were chosen, 19 times out of 20. Of the more than 100 elementary schools in the Toronto Board, almost all were represented in the samples by at least one student. As a result, students' responses at each performance level were observed, and identified from the sample. These responses were coded into criteria for student performances on tasks directly linked to all of the Benchmark lessons. Through these achievement-level standards, the Toronto Board's criteria for classroom learning were to be addressed.

In summary, the Toronto Board wanted to establish standards of achievement that provided system-wide reference points to assist teachers assessing students that ensured that assessment supported and enriched good teaching practices; that enabled educators to relate curriculum objectives to the actual work of students; and that provided more comprehensive information for reporting to parents. To these ends, the Toronto Benchmark Program is an innovative assessment program that has emerged out of the concurrent assessment and curriculum reform.

Despite some relevant cautions about the shift from traditional to performance assessment expressed (Messick, 1994; Lewis, 1996) in the
literature review of this study, the "shift" experienced by the Toronto Board has gathered momentum in Australia, Britain, the US and in many OECD countries (OECD, 1993). This is evidenced by the increasing use of common assessment tasks (CAT's) or standard assessment tasks (SAT's) at the middle and senior years of schooling, requiring both teacher observation and professional judgement in addition to traditional, standardized tests or formal examinations (e.g., Hill, Brown, Rowe & Turner, 1996).

3. The Innovation. The development of the Toronto Benchmark Program is an example of an innovation that was planned to help schools accomplish their instructional, and assessment, goals more effectively. It was also designed to change teaching from a traditional, didactic, textbook-based approach to one in which learning is viewed as a process by which students construct their own knowledge through active participation and hands-on experiences. In order to achieve this goal, the Toronto Board felt that teachers' attitudes, fundamental beliefs, and basic assessment practices needed to coincide with a student-centred approach.

Although reforms have been underway in various school subject areas for a number of years, there is evidence that the changes teachers are making in response to these reform strategies are falling short of reformers' goals (Cohen & Ball, 1990). This section deals with the question of how reformers expect reform strategies to change teaching practices. Historically, different orientations to curriculum represent diverse emphases on what should be taught, for whom, how, and with what resources (Eisner, 1979; Prakash & Waks, 1985; Vallance, 1986). The orientation to behavioral objectives (Bobbitt, 1924; Charters, 1924; Tyler, 1949) and taxonomies (Mager, 1962; Gagné, 1970; Ebel, 1979; Bloom, 1956) designed to organize objectives into
levels of learning can be seen with the introductions of 'common' (i.e., national or provincial) curricula in Great Britain, and Ontario as well as throughout the United States in conjunction with performance-level content standards in almost all school subjects at every grade level. How the participants of the reforms are influenced by the assessment strategies are dealt with in the findings of this study.

The approach to student evaluation represented by the Benchmark Program honors wholeness in several ways. First, most Benchmarks incorporated multiple learning objectives. The Benchmark tasks were designed not only to operationalize several objectives simultaneously, but were also designed to resemble normal, complex, open-ended and life-like classroom activities. Many of these tasks were designed so that students' performances could be observed on videotape and be evaluated holistically. This approach meant that a "whole student" could be observed performing a "whole activity". It also meant that many characteristics of the student as well as the context within which the student performed could be considered. The holistic scoring utilized criteria that were drawn from students' performances across a wide range of thought and process skills. Second, the Benchmark units included introductory and concluding discussions, prompts and questions so that students and teachers could interact with each other during the Benchmark activities. The tasks reflected the social and fluid nature of learning and evaluating. That is, students could learn during evaluation and be evaluated while learning. The Benchmark activities left room for interaction, sharing and negotiation, and for different styles of learning. Third, the Benchmark Program recognized the student and teacher as complex wholes, both of whom could attend to each other and to many other stimuli simultaneously. Both
could create questions, solve problems and make judgments within whole contexts that were continuously changing within their Language Arts classes.

Second, student assessment was expanded to include multiple modes of response, and levels of thought. For example, the Program allowed students, using higher-order thinking skills, to respond through drawing, making models, role-playing, oral discussion and the creation of media. In previous Language Arts programs, students were confined to responding and being assessed on the basis of multiple-choice questions or on holistically scored written responses. While this practice rewarded only a subset of all students who already possessed such skills, it penalized others for whom these response modes were inappropriate or who needed more development, e.g., ESL students. Because the Benchmark units have expanded opportunities for students to express themselves in a multitude of modes, and across levels of thought, the Program recognizes a wider range of student abilities across a larger pool of diverse students. In the process, it uncovered more aspects of student abilities to be developed.

Third, the Benchmark Program integrated rather than abandoned the assessment practices of former Language Arts programs. For example, in completing the tasks, students worked both alone and in interaction with the teachers. Both analytic and holistic scoring were used. Some tasks were close-ended, and some were open-ended. The processes as well as the products of problem solving were evaluated. Students were required to answer orally as well as with paper-and-pencil. Higher-level thinking skills were evaluated as well as the skills of recognition and recall. The Benchmark Program could be used to inform classroom programming as well as to report to parents. The Program identified system standards that satisfy public
demands for measurement as well as permitting the classroom teacher to stay in control of assessment and instruction.

Fourth, a growth rather than a production model of learning pervaded the Program. Most units demonstrated five levels of development for the task-at-hand. Thus, students may develop or grow at a faster rate with respect to some tasks than with others. For example, a grade three student might perform at a level five on one Benchmark task, and at level three on another. Thus, the tasks are thought of as a collection of references for evaluating the growth of individual students. The Program did not put a ceiling on that growth, limit the growth to a narrow range of intellectual activities, or suggest that performance at a lower level means failure. Rather, the Program is developmental in nature, and demonstrated that it is difficult to determine what students at any age or grade level must be learning and achieving.

Fifth, the Benchmark Program illuminated the role of Language across the curriculum. It is through its emphasis on Language that higher-level thinking and process skills could be made observable in students' classroom activities. For the Toronto Benchmark Program in grades three and six, this emphasis has caused a shift in approach from teacher to student-centred Language Arts, i.e., whole language. For the classroom teacher, this shift meant that youngsters acquire language rather than learn through direct teaching; that language learning is child-centred, not teacher-dominated; that language is integrated rather than fragmented; that children learn by talking and doing rather than by passive listening; that they learn to read and write by engaging in experiences with literature and writing, as opposed to drills and workbook exercises; and that children learn best in interactive problem-solving situations rather than in isolated individual tasks.
Sixth, communication, openness and sharing were favored. Samples of students' performances (both videotape and print), provided with the scoring criteria, elucidated the objectives and standards of the educational system. The teaching materials and students' work were designed not as secretive tests but as reference materials to be studied by teachers, parents and students. Thus, collaboration and sharing were critical to success in developing the Benchmark Program.

Seventh, there was an increased emphasis on professionalism that included a greater involvement in the design and development of curriculum materials. Because the Benchmarks are information and not tests, teachers can shape the way they are used in assessing student work, and in reporting to parents. The role of the school principal and teachers is an active one - to experiment and create the best ways to use the Benchmark Program.

Thus, the Toronto Benchmark Program necessitated a shift in the role of the classroom teacher from that of passive transmitter of a ready-made program, to that of active developer of curriculum in keeping with the characteristics of his/her students.
Chapter VII

Findings of the Study

The findings in this section have been organized and reported according to Stake's (1987) framework of "antecedents", "transactions", and "outcomes" as an aid to delineating the assessment-curriculum consequences in each research site. In actual practice, degrees rather than dichotomies commonly exist in the relationship that each teacher displays between his/her assessment and instructional practices. Participants' narratives with regard to the various themes in this chapter will illustrate the location on the continuum that exists between assessment and curriculum.

1. Antecedents - Precursors to Change

The antecedents in this section refer to aspects within each research context that constitute the beginning stages for mobilizing the Toronto Benchmark innovation. Teachers' activities in preparing to use of the Benchmark innovation are highlighted. While some of these activities occurred after the Benchmarks were mandated for official use, almost all preparations in Montrose took place two years before such preparations began in other Board schools. The themes that emerged from the study's data on the antecedents centre on conceptualization, adoption, process, support, communication, staff turnover, teacher resistance, portfolios, and continuing development.

1. Conceptualization at Montrose. When asked "Do you sense that there is a problem regarding understanding or conceptualizing the Benchmark
program?", Montrose administrator Merton's comments shed light on the viability of the organizational (and behavioral) approach taken by the Board to present the Benchmarks to teachers for development in their respective schools.

I think it is hard, and I don't think that the entire teaching department understands it despite four years of working on it. The Board spent a fair amount of time inservicing principals and vice-principals who were then supposed to work with teachers. The Board had a little bit of time working with teachers...they figured that would do. Part of the problem, of course, is that not every principal and vice-principal bought into the [Benchmark] idea as easily or understood as quickly. So some of them just didn't do it...some teachers felt that we who have to work on it, have had less training than office staff who don't have to use it.

The same administrator referred to the "closed door" syndrome that has affected some teachers who were, at first, somewhat reluctant to engage in the development of the innovation. She linked the importance of dealing with the fear of uncertainty, openness, and collaboration with other teachers as a condition for understanding the Benchmark Program.

The real crux of getting over the fear of the Benchmarks...doing Benchmarks...is teamwork amongst teachers. One of the problems is that teachers for the last many years have worked behind closed doors, and no one has seen much of what they've done...even the teachers who have done terrific things...But if teachers can work together...all of a sudden there's trust in someone else...and then you lose the fear of someone else watching you, and you have coordinated efforts...

Unsurprisingly, teachers in Montrose were able to state their understandings of the Benchmark program very coherently and concisely. Each teacher highlighted a particular aspect of the Benchmark Program they valued. For
example, when Muriel was asked for her meaning of the Benchmark Program, her reply was in terms of standards:

Its a system of standards so that all teachers across the system can use the same set of standards when evaluating children's work...so that Mr Jones downtown and I have the same standard against which to evaluate work even though we've never met.

Muriel also highlighted valued aspects of her portfolio system:

Well, I meant it to mean a place to organize work as a very convenient storage container kind of thing...to hold materials together easily so that they were together in one place for writing assessments, doing evaluation, in one place for an interview so that they were easily accessible to parents...you could show a progression. When it got going, it seemed like a good place to have a tracking system as well.

Margaret interpreted the Benchmark and portfolio programs as:

I see the Benchmarks as a kind of natural support for whole language programs, and the portfolio is an ongoing collection of products...things that the children produce in the classroom. There are a couple of different ways of doing it: I have an ongoing portfolio where everything that the kids put on paper goes in to the file, and then from that file I select what I call significant pieces that go into a permanent file that will then follow the child through the school.

Michelle commented on the fit between the Benchmarks and portfolios:

The Benchmark program is the integration of the specific curriculum and the assessment that goes on within it. I see it as a tool for teaching. I found that the Benchmarks provided me a comprehensive package helping me build on what I was already doing. The standards, the criteria and the samples of completed tasks were particularly useful. Portfolios for collecting students' work for grading and for helping them progress go hand-in-hand with Benchmark activities and helps in reporting to parents.
Con~uauization at Arcadia. The Language Arts Coordinator for Arcadia believes that many teachers there see the various initiatives of the Board as separate instead of as altogether, e.g., "anti-racism" and "language across the curriculum".

A teacher recently said to me "You know that anti-racist poetry you got me for African Heritage Month, I don't suppose I could use that to do a Benchmark, could I?" and I said "Sure!" and she said "I could...you mean it can be all together?" So they see them as separate and they're not.

The coordinator added that she felt some teachers in Arcadia saw the Benchmarks as programmed instruction. For example,

"Today I have to do a Benchmark here, so I'm going to do that here" instead of "Well, these are the things that kids need to be able to do, and once they're in my program, and I've worked with them in different ways, then I can do a formal assessment of this based on the criteria, and I can now report that to someone else.

Annette, when asked for her understanding of the Benchmarks and portfolios in her school, replied:

Portfolios are a folder for children's' work which shows their progress with knowledge, skill, attitude, and participation skills through their classroom activities. It also shows the process by which students grow in their academic abilities and levels of performance. And with the Benchmarks, assessment becomes "fine tuned"...the objectives become more clear cut. The Benchmark and portfolio programs show me three things: one, the development in students' learning...that is, how far they've moved from here to there; two, where kids are having difficulty...and three, where I need to modify the program, for instance, instructionally, other-activity oriented...
From Annette's perspective, in order to understand the Benchmark Program, it was necessary to take it one step at a time. She said that if you look at it as a whole, it's just too overwhelming that you might not want to try it:

So I think you have to start small...look at what you're given in the binder...look at the type of activity and skill and see if you can collate some material that would help you to assess that skill over a long period of time...maybe starting with a smaller story...or a more simple story and then the next time you do the same Benchmark...a more sophisticated piece. So its sort of gathering information and then deciding how you're going to use it at what stage in the year to help in the process of helping a child improve in that skill you are assessing.

Adelle's understanding of the Benchmarks and portfolios were aimed more at the range of abilities and interests of students in her class:

I think Benchmarks are a basic descriptor of where kids are at in specific grades...what they can do at different performance levels. Part of it also involves reporting to parents...a very important step if the child stands out from the rest of the class, one way or another.

Adelle expressed concerns for a program becoming standardized, one which would generalize to the child. However, the concern of time is also a factor:

...there are kids who need particular kinds of help, so you're always modifying it. And its time, I just don't have the time to sit down and modify it...read it with all the instructions in a way that would really satisfy and work to advantage for my students.

When asked if she used the portfolios as well, Adelle replied:

I collect samples from the kids work in terms one and two and put them in their Benchmark folders. The work in their folders gives me an indication of their
performance over time...especially for kids that stand out. It also gives me information for Learning Centre support, and for placement of students into special programs.

Arvel from Arcadia had a special education class. She had serious concerns for the "fit" between the activities in the Benchmark Program and her students' levels of abilities...She explained how the content and criteria of the Benchmark tasks are inappropriate for her special needs students.

The content and criteria of the Benchmark tasks are inappropriate for my special needs students. If the kids can't listen, how can they recall the story? Some of my kids can't read...though they are good sketchers...

II. Adoption at Montrose. Comments from teachers regarding Montrose administrator Mathews' method of involving them in the development of the Benchmark innovation at Montrose relate to staff participation:

He [the school principal] essentially "volunteered" everyone to come on board with it. He said to all of us: "For those of you who do not wish to use the whole language approach, the [school transfer] forms are in the top of that filing cabinet". Everyone got into the whole language process of instruction.

Further, this administrator explained his position on the Benchmarks, and how the adoption process progressed in his school,

Everybody is going to take two Benchmarks...this was my mandate...they [teachers] talked about how they were going to do that, and we decided...they decided...that everybody was going to use the same ones. And we were going to use a similar one in primary as we did in junior...so that we'd get a sense of the continuum.
Administrator Mathews concluded:

It just fit in because we were working on the developmental programming aspect and we were having regular meetings...anyway...talking about how kids learn...

Teachers at Montrose believe very strongly that it was a teacher's instructional program tailored to a particular level of student learning that was critical to the entire educational process. They also believed that new programs introduced to classroom teachers need to be kept in perspective. For example, Michelle explained:

When you know what your program is...either on paper or in your head...then you know where you're going, and usually the way to get there. So once a teacher eventually has this part developed, and someone comes along with another program to get started...you can take a look at it...take what you need from it...what will help you...and keep on going.

Muriel clarified this perspective in how she adopted the Benchmark Program:

The Benchmarks are really closely related to my curriculum. Benchmarks are very broad, there's lots of room in them for assessing all different aspects of a kid's progress in language, and so I find that I can take bits and pieces of Benchmarks...there's certain things that I can do easily...there's certain things that adapt well to small groups or individuals, and some things you can do with your whole class.

Margaret added that using a whole language approach to instruction promotes adoption of the Benchmarks:

The program in your classroom really has to complement the Benchmarks, and the Benchmarks will then complement the program. I think if you are running a true whole language program, that becomes the case. If they're running the old fashioned
subject-by-subject curriculum, it doesn't work as easily. The Benchmark program for me fits nicely into the program that I run because it supports what I do, it helps me in assessing the children that I work with, it helps me to evaluate my own program and to evaluate the kinds of programs these kids should be placed in.

Maureen said she looked at the Benchmark innovation, and found she already was teaching it:

I teach them as part of my regular instruction. I have been emphasizing skills like story writing, closure, and grammar for example, since day one.

Adoption at Arcadia. As an indicator of accepting the Benchmark Program, teachers interviewed in Arcadia expressed a variety of terms to describe the ways in which they "accepted" the Toronto Benchmark Program. Critical to understanding the degree to which teachers "took in" the program was the relationship between their (former) instructional programs and the (new) Benchmark Program. The merger of the "former" and the "new" was exemplified by Arcadia teacher Annette as:

I think I look at Benchmarks as fitting in to what I already have. I don't look at it as "Benchmarks...lets start from the Benchmarks and I build my program around it". I think that they're just added aids for teachers. I use Benchmarks to enhance my program, to build skills, to improve my curriculum and my methodology in relationship to all the other things that helps students to learn. My instruction encompasses the Benchmark program...so it [instruction] isn't limited to that suggested by the Benchmarks.
When asked if it becomes another dimension to what she was already doing, Annette's reply was:

Sure...and not to preclude any other...such as analytic assessment or standardized formalized assessment or subjective teacher observation or peer evaluation or self evaluation. It's in conjunction with all of this.

As Annette adopted the Benchmarks as a "program enhancer", i.e., for her own developing program of instruction, Adelle saw them as friendly additions to her struggle in addressing the ability and interest needs of her students:

I take a look at them and and try to match them with my kids...what I think is appropriate for them. If I see some good ideas for them to do, I gather materials and change the task around a bit to suit them.

Alex views the Benchmarks as "motivations for learning", interesting activities that captured his students' attention. Still, to him the Benchmark Program was not the "be-all-and-end-all" by itself. In fact, he stated:

The Benchmarks reinforced what I was already doing. The Benchmarks are a fairer assessment than a standardized test...some of my students never had a [standardized] test before.

Adelle's comment in the Arcadia site related to her first days with the Benchmark program:

The very first day [of teaching] I asked if there was a language document, and they [school administrators] said "No, we don't have a language document, we have a Benchmark". I asked "What are Benchmarks?", and nobody could really answer that. I was so bogged
down, it took me another week before I could get to it...I was so frightened, I felt I had to do everything right...

Alex reminded us of the linkage between the rationale for the program, implied staff development, and its adoption:

I think we just have to look at it as "this is why its positive, this is why we're doing it, this is how we're doing it". I think the how is probably the most important thing: How are you going to use them? And its easy if you say...like standardized testing...I'm going to do L1 today, I'm going to do L2 tomorrow, and I'm going to go through the 24, and then I'll start at L1 again and see how far the child has gone. I really think its the how! The how was never really dealt with in a mature, responsible, positive way. And I think that's the downfall of it at this point.

Though an experienced teacher, Arvel had not been in the classroom for the previous few years. She was in the midst of developing her own instructional program, but was also trying to integrate the Benchmark Program into it. She found she doesn't have the time to do the assimilation, and as a result said: "My instructional program is constantly in a state of flux". The nature of adoption for Arvel could best be described as "piecemeal".

III. Process at Montrose. At the Montrose research site, administrator Mathews explained how, after arriving in that school, he began the process of developing the Benchmark Program:

...one of the first things we did when I arrived as the principal, I did all the sessions. I conducted them, and that was non-threatening...And then I said "Everybody's going to take a turn", and some people were very, very threatened by this. But once they had done it, their confidence level just soared!
Michelle's reaction to "volunteering" their time for Benchmark development was:

At first some of us were a little annoyed with having to attend these meetings every two weeks, but as time went on...we looked forward to them....it gave us all a chance to really ask questions about the program...to find out more about it...and to help each other use it in our classrooms...So, our classroom instruction went along very smoothly...

Margaret went on to say that administrator Mathews was supportive in helping them build confidence in using the innovation in their classrooms:

Every two weeks during our noon lunch hours, he held professional development sessions in whole language instruction for our staff during that first year [of development]. We soon realized we could handle the program!

Mr. Mathews explained how his staff had incubated on the idea of the Benchmark Program for two years. During this time, teachers and administrators in Montrose focussed on the purpose of the Benchmarks, created materials for classroom use, collaborated on the holistic assessment procedures and reporting , and gradually came to understand what the program was all about through a series of meetings. In the next two of the entire four years of developing the Benchmarks, Montrose teachers and administrators were able to practise, and fine tune the program through their own continued classroom practice and professional collaboration. Mr. Mathews commented on this experience:

Now, one of the things we said was we worked at it for two years just going through the Benchmarks as a library...learning what was in it...including those kinds of practices in the program if they weren't already there.
Administrator Mathews then included parents in the development phase:

And then we said, "Okay, it's time to bite the bullet. We're going to use them as evaluative tools and we're going to use them to report to parents". Now we already had three parent evening sessions with the parents. We had made sure teachers were already acquainted with the Benchmarks before we had any meetings with the parents...and they were highly successful...

**Process at Arcadia.** When asked to identify the major issues in the development of the Benchmark Program in the Arcadia research site, administrator Ashton recalled the experiences of her teachers for the three school levels. At the primary level:

...there was really no obstacle because those are the things that teachers have been regularly practising. And so for them its not new, but I think its helped them in their evaluation and in looking at what they're doing in the classroom...their own performance and also the performance of their students. So its helped them in the delivery of their programs.

She added [for teachers at the junior level in her school]:

What I see in some cases...that it was viewed as something external by those teachers. Because it was mandated for them...they have to do such and such...that now they have to look at presenting the program in a different format...but there is still a long way to go.

However, at the senior level where there has been a tradition of formal assessment, Ms. Ashton commented on Benchmark development amongst three teachers of varying years of teaching experience. She suggested that for those who have become entrenched in their teaching, it may be more difficult for these teachers to depart from their established practice:
...in some cases it has been difficult because not all of the staff have been very open to experiment. One [senior teacher] is open to it and has been trying to do the Benchmarks...because of being very new [to the Benchmarks]...and because he sees them as a core program as opposed to a specialty. But the other two senior teachers who've been in the business for a long time...they...may treat them as a separate entity more than integrate it into their classroom programs.

Arcadia's administrator went on to describe the "isolationist-collaborative" continuum concerning the way her staff members work. She cited the importance of having a common vision as a base for staff sharing in curriculum development:

The problem has been that two of them think the same way...they all work on their own...in isolation...they never work as a team, and the other doesn't participate because she doesn't share the same philosophy [as the others]. Whereas, if you look at the kindergarten to grade four levels, they share the same vision...and tend to work together as a team.

iv. Support at Montrose. Muriel in Montrose reflected on how the professional development activities in her school provided support with the Benchmark and Portfolio Programs:

Yes, it's an ongoing thing throughout the year...When we were first working with Mr. Mathews, he had us working as a staff, talking about Benchmarks, looking at Benchmarks...the videos and print-materials and not actually using them...just getting comfortable with them and familiar with them...and I really think that's extremely important. That made us, when we actually decided to get involved, use them and be comfortable with them and understood them.
When asked if Montrose involved the community in their network of support, Margaret replied:

Yes...we have parent-teacher meetings every month. Those are executive meetings, not big meetings, but every year we have a huge thrust with a focus on curriculum...I think there might be two this year...one of them being Benchmarks...again because we get new parents in the neighborhood all the time.

Michelle also referred to her upcoming team meeting:

We're meeting tomorrow night...the 3's and 4's...we're all bringing examples of a common Benchmark that we used...and we're just going to look at how we holistically marked them and make sure we're in agreement and talk about whether we had any problems and do we want to do a common one again before June, and if so, which one? So just sort of organize them...

Margaret reflected on how the professional development activities in Montrose focussed on preparation for use:

We've had professional development days where we've devoted the entire time to meetings where we've cross-grouped everybody from kindergarten to grade six teaching staff mixed together, and we've looked at certain Benchmarks that look at measurement in grade three, again at six, again at grade eight...And we've sat down as teams of people across the grades to look at these Benchmarks and see the development of skills that kids are expected to pick up, and then we've gotten back together as a whole staff and discussed...So we've spent a lot of time on professional development.

Support at Arcadia. Arvel commented on earlier plans for the four junior-level teachers to get together on their own, independent from the school in order to provide help for each other in developing the Benchmarks. Unfortunately, their plans did not materialize. She recalled:
These meetings weren't built into the system...they weren't a regular part of our school duties...Too many duties got in the way, and we didn't meet too often...they [meetings] just fell by the wayside.

Arvel linked her limited opportunities for Benchmark development to her desire for staff development functions. When asked to comment on the need for professional development with the Benchmarks, Adelle also commented on the inadequate support received through professional development:

We just don't have enough P.D. days...you know we had only one of our days devoted to holistic marking. Well, what can you do...teachers are tired, someone lectures at you...they [teachers] just slouch back and turn off. So, some teachers have just given up on Benchmarks...hoping it'll blow over in a couple of years...others are doing only bits and pieces of it.

According to the administrator, the professional development opportunities given to teachers in Arcadia as support for developing the Benchmarks program were through Board rather than school initiative. Through the Board's Benchmark Steering Committee, the staff of Arcadia actually did a Benchmark together. Annette expressed great satisfaction with first-hand experiences in having worked through the Benchmark activities in such a manner:

When we discussed with each other how...what we saw in the short reading and why we scored it the way we did...after working through this, we found this really wasn't so bad at all, and in so many ways it was a big improvement [Benchmark and holistic scoring] because it showed us in so many more ways [other than analytic scoring] where students needed help and where we could spend time working with them.
But, this session in Arcadia took place only once. Adelle in Arcadia identified the critical link between professional development and development of the Program as that of confidence:

Oh...its a really major concern...I mean we get better...we develop with practice. But without practice and discussion a lot of teachers are just shelving it...they don't know how and aren't confident with even trying them out in their classrooms with their kids.

Still, Arvel from Arcadia cited positive experiences in receiving help from her more experienced peers. Though time was a problem, staff support helped in her development of the Benchmark Program:

The staff here has been very supportive, especially the experienced ones. I share a lot of ideas and resources with other grade 3/4 teachers here during lunch and on my own time. We have no scheduled time for these meetings, and learning [the program] at the same time as teaching it...its just extremely overwhelming!

As well, Annette spoke positively of the help she received from a professional development session on holistic scoring:

It takes a lot of time, and at first I was afraid to do it for fear of being biased...too subjective...But, during a PD day, we worked through one [Benchmark]. We shared ideas about the different ways we did the same Benchmark. Scoring it was something else! We learned of each other's approaches to what we saw in the student's work. After that meeting, it was easier to score...and use the Benchmarks in different ways in the classroom. I went ahead and used the Benchmark in my class..I don't think I would have done so otherwise.

v. Communication at Montrose. The last teacher (Maureen) interviewed at Montrose recollected the professional support she received and her assessment
of it in terms of communication while using the Benchmark Program with her students:

We have team meetings...the five of us meet probably on average once a month...sometimes more often. And we do sit down and talk about what we're doing, we plan units together, we do share materials and all our resources...ask each other questions. So communication with other teachers has improved.

Maureen also referred to this communication as a regular part of her school life:

In fact now if we don't meet for a month the teachers in my team especially will say "We really need to get together and talk about what we're doing". Keeping in touch has helped us tremendously.

Administrator Mathews from Montrose recalled the essence of his communication to his staff:

Now, in a way...I mean the Benchmark Program is a very compound...a very complex kind of thing and it is both program and evaluation...they go hand in hand. And that's why I say that the beauty of the Benchmark Program is both program and evaluation...because you've got to inject that rich program in in order to evaluate all those aspects. The two go hand-in-hand.

Communication at Arcadia. In Arcadia, data relating to support for change, and teacher resistance addressed the issue, i.e., the lack of, communication amongst staff and administration in developing the Benchmark Program. Major indicators of this characteristic in Arcadia were that the source of professional support for program development was Board rather than school-based; school time was not officially scheduled for teachers to
meet to plan, organize, and problem-solve for Benchmark development. Some teachers simply refused to participate in the entire development process. Although staff turnover in Arcadia was not an issue, communication amongst staff was limited to conversations on the Benchmarks that took place irregularly, and that were based on teachers' own social networks.

vi. Staff Turnover. The issue of staff turnover applied to Montrose, but was not significant to the development of the Benchmark Program in Arcadia. Since administrator Mathews from Montrose has retired from the education field, administration and staff new to Montrose [and to the Benchmarks] have arrived on site at the time data were being collected for this study. This change has affected the development of the Benchmark Program in Montrose. However, Maureen provided an example of the current situation for a number of new staff in Montrose:

I found just taking things from this separate binder...I mean...it's [Benchmarks] not part of my program...it's on the shelf at the side of the materials I use for my program...I found them unwieldy...like I was doing them just to do them. So, I just put it down...and don't really use it at all.

Maureen's remarks pointed strongly to the need for staff development to be continuous, ongoing, and systematic. There needed to be opportunities for teachers to try and use the program, talk to colleagues about it, and integrate it into their [instructional] program.

vii. Teacher Resistance. Teacher resistance to the development of the Benchmark Program was a concern for Arcadia, but not for Montrose. When asked about any reluctance to develop the Benchmark program, Arcadia
teachers Michelle and Maureen identified gaps in development, and reasons for teacher resistance. Michelle commented:

...a lot of us were resistant. Teachers who saw themselves as speciality teachers thought the Board would make us do all the groundwork for ourselves in each of our subject areas. And when we'd say "That's a good idea", they would say "You do it". But if it didn't work out, guess who would be held responsible!

Michelle also cited a host of other conditions she thought helped contribute to teacher resistance to Benchmark development in Arcadia:

Time to do the Benchmarks, to prepare for them is problematic. It is hard to change after teaching one way for a long time...some are very set in their ways. There are also difficulties coming from the teaching job itself...large classes...single and needy parents, and the recession means many kids don't get enough support for their growth.

When asked if she met with other teachers to work through the Benchmarks before using them, Maureen replied:

I would like more opportunity to get together with the people at the junior-school level and discuss it. There are at least two or three teachers I could do that with, and we just haven't had the opportunity or the time for it.

Maureen identified two staff members at the junior level who were actively resisting the Benchmarks development, and who previously were not identified by the administrator:

There were six staff members at the junior level, two of whom refused to come to meetings...who refused to cooperate with others, refused to do the Benchmarks...
viii. Portfolios at Montrose. Both administrator Mathews and Margaret from Montrose commented on the development of a system designed to record student data connected to Benchmark activities, and to communicate these results meaningfully to parents. Mr. Mathews commented on the development of the portfolio system in Montrose:

[Muriel] started playing around with the chart, and then changed it to the folder. And it was interesting to watch the other teachers because they were doing things as well, it wasn't just [Muriel]. One had a graph, one had a circle...where she'd put in charts. Once [Muriel] folder was developed, we thought "how beautifully it integrated everything...visually for teachers, and how simple it was once you had gone through it".

Muriel identified time as a motivating force behind her efforts to find a useable system for recording student performance:

What was really time-consuming before the portfolio was finding a method of recording all this so you could use it for interviews, for reporting to parents, and in report cards. So there had to be some way to save time...so I just tried to think of the quickest and fastest way to organize all that information and make it meaningful and handy and professional looking for interview purposes.

Administrator Mathews went on to explain how the portfolio system fit into the implementation at Montrose:

It [portfolio] was the missing piece of the puzzle that linked our classroom instruction with communication of students' results to parents! So, all the teachers said "Hey, that's for me" and once that happened I went to the associate director he said..."This is exactly what we hoped would happen...this is better than we could have envisioned because it came right out of the classroom!"

Muriel explained that:
From my classroom it went to the Area North Board where it was piloted, and that started last February (1993) and it was then presented to the Toronto Board and it went to city-wide pilot which includes every student in the Board...

Muriel also expressed the close relationship between assessment, instruction and the portfolio system:

I think here that we have it [assessment] integrated in to our programs. They don't see it as we report to parents and then we have a portfolio...its altogether...I don't think that's the case in any other school in the City. I think for a lot people its still an add-on...its still the Benchmarks and the portfolio - just one more thing to do... and part of that is ownership because we feel that sort of belongs to us...and part of it is just being familiar with it...you know, that it doesn't really take any more time...

Since 1988, Montrose has combined the portfolio they developed with the Board's Benchmark Program.

Portfolios at Arcadia. Data from interviews and class observations concerning portfolios from the Arcadia site were extremely limited. According to administrator Ashton and teacher Annette in Arcadia, at the time of this data collection, there were no school-based, and coordinated efforts for development of portfolios. However, teachers observed had varied, but limited methods for storing, and using the results of students' performances. For example, teachers Annette and Adelle used "student bins" for keeping their Language Arts assignments organized while teacher Alex used individual file folders designed to organize and show each student's development of writing skills over time. Ms. Ashton stressed that in the following year, initiatives would be in place to begin portfolio development for use in teachers' classrooms.
ix. Continuing Development at Montrose. The comments from teachers in Montrose reflected, on the whole, experiences more conducive to having the Benchmark initiative expanded to other courses in their school. Montrose's administrator Ms. Merton was concerned with extending the Benchmark Program well beyond her Board's initial intention:

The next one we want to look at is to see if we can pull in science in to this...it makes it more difficult because there are not science Benchmarks so we don't have Benchmarks correlation between this [science] Benchmark and this assessment, but we're trying to look at the type of curriculum that leads to the way we get our best teaching and we transfer Language to math, to science...

The same administrator viewed the situation as an opportunity to support others:

The other thing is, a school like this where it really is working, we have to help the other schools...[others] haven't come as far along as we have. So teachers like...[Muriel and Margaret]...have assisted in their development with the program. Somehow its easier when it comes from another teacher.

Continuing Development at Arcadia. When Arcadia's administrator Ms. Ashton was asked to comment on her plans for working with the Benchmarks with her staff in future years, she commented:

I think that the next step is challenges for the teaching staff to work towards integrating what the Ministry says with what the Board is coming out with in their Community Reviews...the challenge is to take the Benchmarks and see how they dive-tail in to the expectations of the Ministry and the new Common Curriculum.
Out of the nine themes identified in this section, all but three pertained to both schools. Teacher resistance was mostly particular to Arcadia while the two themes of "portfolio" and "staff turnover" were most relevant to Montrose.

**Summary.** While the two research sites in this study were both in the process of developing the Toronto Benchmark Program, the nature of their experiences associated with developing the program differed markedly. In order to see how these similarities and differences manifested themselves, the data on classroom instruction and student assessment is presented in order to show the nature of use of the Benchmark Program in each research site.

2. **Transactions - Assessment and Curriculum in Operation.**

The transactions refer to "the actual use of the innovation" (Fullan & Pomfret, 1975) in the classroom setting. In order to see how the background and development histories of Montrose and Arcadia research sites have manifested themselves with regard to assessment and curriculum, data on the relationship between the two are presented for analysis and discussion. In each case, the classroom practices are viewed in connection to the Language Arts curriculum in use at each research site. The data used to identify the nature of assessment practices at Montrose and Arcadia research sites evolved from the researcher's observations of classes in progress, and from his interviews with the designated students and teachers.

i. **Montrose.** Comments by teachers in Montrose reflected a balanced relationship between assessment and curriculum in their Language Arts program. Teacher Muriel at Montrose demonstrated the manner in which she integrated assessment and curriculum activities in the lesson "Chin Chiang and
the Dragon's Dance". This Benchmark lesson was developed to assess students' ability to read part of a narrative and continue the story in writing. Students read the first three pages of the story in which a problem was posed for the main character. Students then completed the story in writing. Before the activity, students were asked questions about their participation in concerts and knowledge about Chinese dragons. After reading the first three pages of the story silently, they were asked to finish writing it.

Muriel used overhead transparencies to introduce her grade three class to the expectations for the lesson. First, she showed them a level-two example of a student's assignment of continuing to write an unfinished story:

T: Let's take a look at the spelling...can anyone find a misspelled word here?  
S1: I know one..."maybe" is wrong ("mabe")...it should be "m-a-y-b-e".  
T: Can we tell who Chin Chiang is in this story?  
S2: He hardly ever speaks.  
T: So when we write we should try to have the people in our story say things. What are some things we can have Chin Chiang say?  
S2: He could say that he wants to be in the dragon dance...  
S3: He could ask his friend to be in the dragon dance too...  
T: What happens at the end of this story?...Who knows...(Students read, but do not answer...). Perhaps when we finish our stories about Chin Chiang...what are some things we can say at the end?  
S4: He could say to his grandfather 'I didn't do the dance with you...it was my friend who really did dance with you.'  
S3: I know...'His mama was watching Chin Chiang and his grandfather do the dragon's dance and she felt proud of Chin Chiang.'

Muriel then used an overhead showing the same assignment done at a level-three performance:

T: Let's break it up into content. What makes it good information? How about such words as dragon, dance, and grandfather?.  
S1: They tell us who makes up the story...who we read about...  
T: What about the introduction?  
S2: It tells us what it's about.  
T: What about the story? Is it complete?  
S3: It doesn't have the date on it.  
T: What about the spelling?
S2: Its better than the last one.
T: The paragraphs - how many are there? What about the ending?
S1: Its much better than the last one.

Finally, Muriel then showed her class an overhead of what her assignment would look like at a level-five:

T: What makes this one better than the others?
S4: It has more things in it to tell us.
S5: It tells us a lot of different things that happened.
S2: You can understand what they wrote.
S6: It had a lot of detail in it.
T: Did they put themselves in the story?
S6: Yeh, they tried.
T: So, yes...they put a little of their own voices in the story.
T: Now spelling
S2: Okay, good
T: What about the language...is it appropriate?
S2: It said "the mud soap cleaned the people good". It should have said "the mud soap was used to clean themselves".
T: Excellent.

The students from Muriel's class were then prompted to do their own assignments for the lesson. After twenty minutes of working on their written assignments alone and in groups, the students sat in front of the room while the teacher related the writing assignment to the performance criteria, i.e., amount of detail, different examples, spelling, word use, punctuation, paragraphs, and characters' voices in story.

Student (3) commented:

S3: I wish we could've made up questions instead, this was hard.
T: Remember the first time we tried to make up questions on our own?
S3: Oh yeh, not so good.
T: But, after we tried it for the third time, it got really easy, and we got really good at it.
S7: Yeh, that was fun.
This was the first time for this Benchmark lesson with these students, but throughout the year students did this Benchmark three times. Although Muriel said she'd use different [reading] materials each time, she applied the same performance criteria associated with this lesson to future Benchmark lessons. Another teacher that was observed from Montrose who had already completed this Benchmark twice in the year reported that her students provided more content, showed improvement with word usage, and connected more to the meaning of the story they were retelling. Students' assignments were kept in each of their Benchmark portfolios for handy reference in order to see how they had progressed over time with regard to their lesson's performance criteria, for purposes of showing their assignments to parents during interviews, and also for purposes of grading on report cards. In a subsequent class, each student conferenced with the teacher for ten minutes on the results of their written story telling. After receiving feedback, students discussed ways with the teacher on improving their writing. Students then returned to their desks to draft another version of their assignment.

In Margaret's grade three class students' abilities in reading exposition and retelling it in writing were assessed. Students read a science book about camouflage *How to Hide a Butterfly & Other Insects* (Heller, 1985) and then wrote about it. Before the activity, students were asked questions about camouflage and were rated for prior knowledge and experience. Students were then asked to read the book silently and to write as much as they could about it in their own words. After reading their stories to the rest of the class, Margaret first commented:

T: You said "I" in the story. Were you in the story?
S1: No.
T: Oh, that was a bit confusing for me. I thought maybe you were in the story.
T: Okay. So maybe we can use pronouns here—words that can stand for "I". What are some of these words?
S2: We, they, them...
T: Good. So, if instead of saying "I", you can say 'they' when you are referring to Katherine and her friend. That would make the story easier to follow for everyone.

[Another student reads her story]

S3: (other questions). I don't get the story. Could you explain it?
Reader: Okay, I'll put that in for my next story.
S2: [Reads his story]. I went to China. There are many people there.
S2: What did you do there?
T: Peter, think about what you did there. Then add.

As a result, students took with them valuable information designed to improve their stories.

Michelle summarized the previous lesson, but did not engage students in this part of the class. She proceeded to introduce Benchmark lesson L3-11 (Narrative - Continue the Story) to her students. Michelle introduced "The Tunnel" (by Anthony Brown) - a book recommended by the library for use with the Benchmark program. Michelle announced "We will read it together, and use our great imaginations to make an ending for the story". While reading the story for 20 minutes [with clarity, presence and in varying tones appropriate to the story's plot], her students asked the odd question during the exercise.

T: Now, you are going to hear the end of the story after...
S: Oh...[anxious to proceed with the reading and find out the ending]...
T: But first you are going to make your own ending. Now, what are some possibilities?
   (Teacher goes to the board and takes suggestions from students and writes them down).
S1: What the girl does...
S2: How she rescues her brother...
S3: What happens to the brother...
S4: Time they get home...
S5: How the brother turned into stone...
Michelle explained to her students the point of the exercise:

The reason we are doing this is to get our minds working, and to get practice in writing it down. We are learning how to think further.

Michelle gives her students the expectations for performance:

I expect you to put some thought into this, I want you to put down at least five sentences.

Students went to their desks and individually, began writing in their scribblers. Michelle wrote down the title of the story on the board for her students. She then circulated and helps her students.

T: Don't worry about spelling your words incorrectly...just get your ideas down for now please.
S: [After twenty minutes]: I'm finished. What do I have to do now?
T: [Announces to the rest of the class]: If you are done, please take a book from the front of the room and read quietly until the others are finished.
T: [explains to me]
This is the third time for this exercise. This is the longest these kids have worked so far on any one task. I have to report to their parents, so I gradually give them exercises to get them used to doing them [the Benchmarks]...rather than just "cold turkey".

Students put their assignments in their writing bins, and students switched to another class without using the remaining time to assess their work.

Maureen's kindergarten Language Arts class assessed students' abilities in reading and creating questions to the story presented. Students' questions were scored holistically. Using the picture on the cover, students were asked to predict the content of the story Over the Swamp. The activity began with the teacher reading the first part of the book. The students were then asked to read the rest of the book, and then, based on the story, made up at least four questions which could be asked of a friend. After twenty minutes of working
on the activity, students grouped together and asked each other the questions they had made up on the story. The teacher then summarized the point of the lesson before the bell rang to end this class.

The teachers of the above lessons from the Montrose research site commented on the relationship between assessment and curriculum in their classes. Muriel focused on the importance of the objectives in developing a curriculum, and on the use of assessment to see if the program objectives were being delivered:

The Ministry and Board first set down the objectives of the program. I have to do activities that deal with those objectives, and then find methods of evaluation that will make sure I have fulfilled the objectives in the Ministry and Board guidelines...so its rather a triangle driven by whatever the Ministry and Board are setting up for those objectives. You then set up a curriculum you think is going to deliver those objectives. You evaluate it, and if you think its not delivering those objectives, then you go back and change your curriculum.

Margaret from Montrose explained how her involvement with the Toronto Benchmark Program resulted in increased knowledge and understanding of assessment techniques:

My knowledge and understanding of assessment started at "ground zero" with regard to holistic methods. I began using the Benchmarks dissatisfied with testing and grading as I knew it, but without the tools to do anything about it. Now two years later, I'm more aware of available choices, better able to design assessments that connect to the curriculum that I teach...
The Montrose classroom transactions with the Toronto Benchmark Program have not only formalized, but have also integrated the distinctions between content and performance standards as indicated by Margaret's comment:

You say "Okay", if I want to do this question technique where the kids make up the questions or if I want to have the kids read silently and interpret it with a picture, I get those things prepared ahead of time...those are the things I want to do...and then as I'm working through my unit I decide "Where can I use that questioning technique in this unit?", "Which story best lends itself to that assessment tool"?, "Which poem might best lend itself to the reading silently and illustrating as another assessment tool?"...So as I'm planning a unit, I'm already thinking about the Benchmarks assessment tool I'm going to use in it.

Michelle in Montrose commented on the relationship between assessment and curriculum in order to develop students' skills:

They both [assessment and curriculum] complement each other...so the materials development and the assessment gives you information regarding what level of materials you need in order to develop what skills for each student. I see both meshing together...with a very close fit if you have a well developed instructional program.

Maureen made a comment about how she saw the relationship between assessment and curriculum:

I've been doing it backwards. I thought curriculum came first followed by an assessment which was administered based on what I covered in class. Now I know that it is better to concentrate on broad goals, and then select a curriculum that will give students the knowledge and skills necessary to get there.

As well, all parents of students from Montrose expressed a desire to have an equitable and complementary relationship between assessment and instruction. The following is representative of the comments regarding the
assessment-curriculum relationship by parents at Montrose reaching all students:

I prefer to see programs that come to some equitable agreement between testing and instruction...as opposed to have a program for my children that relies on constant testing...because, unfortunately, tests don't always reflect what one already knows...And those students who would have tended to have done well would get through anyway...so the tests aren't telling us anything we don't already know in a lot of cases...its the other ones that slip through that we have to figure out how are they reached, and how are they assisted with their learning.

A comment from another parent at Montrose emphasizes the ongoing nature of performance:

I think that's really important to bridge the two...because first of all, testing puts undue stress on everybody, and anybody can cram for an exam and forget it. And actually, if you think about how kids learn how to swim, its progressive and you have benchmarks...you know its exactly the same thing...for your red badge you have to achieve this, this, and this...and its progressive, so there's no big testing thing at the end of the session...its ongoing and when the child has acquired the skill, its ticked off on the sheet, and then when all of the skills have been accumulated, they move on to the next level...

Montrose's classroom transactions treat content and performance standards as two interdependent and necessary components of an assessment and curriculum delivery system.

ii. Arcadia. In the Arcadia research site, the content and performance components of the classroom transactions remained relatively separate. Alternatively, teachers from Arcadia viewed the assessment-curriculum
relationship within their Language Arts programs as unchanged, i.e., as remaining assessment-driven.

For example, Annette from Arcadia introduced the lesson on the Olympics to her twenty students. Her students brought in newspaper clippings, magazines and artifacts on the Olympics, and put them in envelopes for each student. Her class discussion on the Olympics touched on the meaning of inter-departmental mail; the identification of some Olympic activities; clothes worn by the Canadian team; and on the Olympics being environmentally-friendly, e.g., use of the color green in advertising, and recyclable containers for refuse. Her students then chose to work in groups of four on a project in which they created artifacts and material relevant to some aspect of the Olympics. After they had handed in their projects for marking, Annette discussed each project and left them at the front of the room for public display.

Adelle introduced her unit of study by handing each student a copy of the fifteen page booklet (Individual Novel Study) which contained directions on how to complete the assignment, and a bibliography of novels students could use. Students were given a list of twenty-two activities. Of these activities the first three were compulsory:

1). Vocabulary - Choose 25 words from your novel. Give the meaning and use in a creative sentence. Divide each word into syllables and show where the stress falls.
2). Divide your book into 3 parts and write a summary of each part. (75-100 words).
3). Make a list of three questions and answers for each of the 3 parts.

Examples of the remaining nineteen activities were: "Write a different final chapter"; "Create an interview with a main character"; "Create a crossword puzzle based on the book"; "Design a book jacket"; "Make a shadow box of a
setting in your story"; and "An activity of your own. Be resourceful and creative". Adele from Arcadia followed a different strategy from Muriel in Montrose. Using assignments created by previous students, Adele showed and described examples of the completed tasks to her entire class, picking out particular aspects, i.e., introduction, purpose, presentation, costumes, questions, the letter, book jacket, book mark, interview. She also discussed ways of answering fully, expressively, and interestingly. Continuing, Adele cautioned:

You must look up the meaning in the dictionary. Vocabulary, we got to get this right. We've been at this for a long time. Start sentences with a capital letter. I will be very severe in my marking. You know how I mark your work, I look at what you can do...

She included a sample assignment completed by an ESL student she had:

The language needs improvement, but the thoughtfulness of the characters really stands out...

Near the end of the class, Adele handed back her students' corrected assignments. Students were given handouts explaining their next [novel assignments] and worked on them until noon. In the following classes observed, students worked on the activities they chose by themselves. While a minority worked in groups of three or four, Adele circulated amongst the students in order to guide their activities and offer her support where needed.

Alex from Arcadia used the Benchmarks on two occasions during the year. Both times the objective was to create endings and texts to poems. He stated that what really mattered to him in creating a successful lesson was providing
students with the appropriate reading and resource material. Alex felt this was no extra burden as he was in the habit of constantly changing his curriculum to suit the needs and interests of his particular students. He used writing portfolios with his students in grade seven Language Arts class. Each student had a writing folder, and each had a table of contents with the date each entry was completed and edited. Emphasis was given to editing writing difficulties, incomplete sentences, capitalization, punctuation, and word usage. By the end of February, 1994, each student had 16 entries in their writing portfolios. Alex used these folders with parents: the table of contents showed parents the assignments their children have completed, have not completed, and the actual performance results. He adapted the Benchmarks to his own program as he saw fit. The Benchmarks, he thought, reinforced what he was already doing. Alex also believed the Benchmarks assessment to be fairer to students than was a standardized test. He says the Benchmarks have helped parents to help their children with their reading activities at home, to monitor their TV viewing, and their game time. Some of the parents took down the program's criteria of performance in order to help their children improve. (However, out of 30 parents, fewer than 10% did so from his grade seven Language Arts class.)

When Arvel from Arcadia was asked what her experience with the Benchmarks was, she replied:

My concern with this is with LD kids, I have to use the junior-school profile...because I've got a junior-school class...but I've got 12 year olds that are at a primary or pre-primary level...which is very frustrating. But by the same token, if the assignment says they must read it themselves and then retell the story, I have to read the story to them...or parts of it to them...
When asked how she would like to see the Benchmarks changed in order to provide maximum benefit to her and her students, Arvel replied:

Well, I would like to be able to use the primary criteria for my primary kids, who are working at the primary level, and the junior level for the kids who are working at the junior level or some kind of mix and match. Now that might be the result of me not having had the training that everyone else has had...there might be some flexibility in that which I'm not aware of. But that's a frustration that I have where some of the stories that were chosen for this age level are far too difficult for my kids...for their listening and reading ability.

Arvel commented that while the library resources for use with the Benchmarks are useable, the time needed for collecting and organizing materials to be used with the Benchmarks could be quite onerous.

Yes, and I find the difficulty is picking..."How does this fit into my own program?" Maybe I'm doing it backwards, maybe I should be saying "I'll teach this first, and then give them practice at it and then test it".

Arvel looked at the level her students work at, and taught them from that level. On the Benchmarks, she commented that:

They help me know if I can integrate material"; and "The program confirms what I already knew. They develop confidence in myself and my students because they deal with specifics...they help determine the performance levels we are at, and allow us to develop from there. They are growth oriented".

Using four Benchmarks with her students this year, she found that her instruction with them, and her students' performance on them got better with time. As a motivator, Arvel constantly adapted to the particularities of her
students abilities and interests, i.e., "...find the trick to motivate my students to do the tasks".

Teachers from Arcadia commented on their classroom experiences in regard to the relationship between assessment and curriculum. Annette from Arcadia commented that she "fits" the Benchmarks into her lessons where she can, but that the Benchmark assessment had no effect on her instruction:

I basically fit the Benchmarks in to where I can, and relate them to the topic I'm working on. So...the Benchmark assessment doesn't alter the instruction I'm trying to develop ....other than things grammatical, basic writing skills...

However, not all teachers felt they could go that far. For Arcadia teacher Adelle, maintaining her own "comfort level" meant keeping the Benchmarks separate from her own instruction:

I've never felt that I could just 'slip in' a Benchmark lesson with my instruction. It would just be too much...you're looking at things like editing...it would be nice if you could just do one for the fun of it. As it stands now, they're separate... just incidental to my instruction. That's my own comfort level at present!

Alex identified time as an obstacle in integrating the Benchmarks with his own instructional program:

I'm trying to develop my own instructional program, and trying to combine it with the Benchmark program. But, I don't have the time to do all this...it's like I have to teach in the meantime...

Arvel recounted her attempt to "include" the Benchmarks in her instruction, and also provide her students with an opportunity to learn in her class:
What I do in the class is not limited to the Benchmarks, but it includes them and goes further...its interdisciplinary, and its language-based...that's very important for me. For example, I start out with a theme...say "television"...I organize tasks and activities that allow the students to practise developing the skills specified by the particular Benchmark objectives. So by having my students create a television guide, they get practice in writing, organizing, planning through the medium of writing...and the specific Benchmark objectives are addressed...

These views point to the varying degrees of "separateness" regarding assessment and curriculum use by teachers in the early phase of Benchmark development. In addition, there was an absence of content and performance standards in daily use. They also pointed to limited movement from a "standardized" basis for instruction, i.e., static Language Arts lesson plans, to a more diverse learning-based instruction with varied learning materials used to support activities that could be assessed against achievement-level standards. However, Arcadia's comments from parents on the relationship between assessment and curriculum pointed to a desire for more cooperation between the two:

Instruction and assessment should be integrated, continuous...because this gives feedback on the program, the instruction, and student performance on a daily basis. This benefits the teacher, student and parent because knowledge of student performance is kept current with the instruction they receive.

Another parent's comment highlighted the importance of feedback in the assessment-curriculum relationship:

I prefer where teachers teach something, then give students a little test and see...a little feedback to see if students understand or not. Not just for ranking, but also for understanding. So the feedback aspect is very important. It helps the teacher to know where the student needs extra help...it is a tool that
can be used in the academic development of the student.

Administrator Ashton from Arcadia comments on the perceived shift towards value on assessment:

Now there seems to be more of an emphasis on the assessment part, and so what you have is the whole spectrum of teachers who have always done it and used it as the basis of assessing not only child performance, but also their classroom program to those teachers who view it as 'I've got to do two Benchmarks...so I'll do them, mark them, and get them over with'. Somehow we've got to help those teachers who only see them as an extra part of their instruction.

Summary. Looking at the literature on formative and summative assessment, and on teachers' instructional methodologies might lead one to view dichotomously the variables associated with assessment and instruction, e.g. assessment as either formative or summative, and instruction as either teacher or student-centred. However, it is important to note that in this study the assessment and instructional practices in the Montrose and Arcadia research sites were viewed more in terms of their maturity of development rather than in terms of discrete and polemic comparisons to each other.

At Montrose, the main feature of the assessment-curriculum relationship was the use of performance-level standards in classroom instruction and assessment activities. With performance-level standards for the Grade 3 Language Arts program already in place, the emphasis in Montrose's relationship was on the students in relation to the curriculum, i.e., teachers were concerned primarily with developing modes of instruction and materials that would provide classroom activities relevant to the performances assessed. Teacher assessment at Montrose was divergent, i.e., it was designed to find out what the child knows or can do with respect to a particular task. This view of
teacher assessment could be said to resemble contemporary theories of learning and accept the complexity of formative assessment.

The main feature underlying the assessment-curriculum relationship in Arcadia was the presentation of data, i.e., subject content, to the students. Thus, the role of assessment was to identify how much of the content had been mastered by the students. Teacher assessment at Arcadia can be said to be convergent, i.e., designed to find out if the student knows or can do a predetermined task.

The differences in emphases in the assessment-curriculum relationships between Montrose and Arcadia school sites were critical in examining the consequences on classroom practices. These consequences evolved from an assessment-curriculum relationship in Montrose that tended to be standards-based - the assessment and instructional practices were focussed on the benchmarks, i.e., standards within the Toronto Benchmark innovation. By comparison, the assessment-curriculum relationship in the Arcadia research site more closely resembled one that was content-based - the assessment and instructional practices, in comparison to the Montrose site, were more focussed on the Language Arts course content as mandated in curriculum and/or teachers' own lesson plans.

III. Outcomes - Measures of Performance

The consequences of the standards and content-driven assessment-curriculum relationships on classroom practices are expressed in terms of outcomes in this section of the study. As with the previous section of this chapter, the outcomes are treated as discrete entities, although in actual classroom practice they may be considerably more interconnected with the lessons' goals, activities and assessments. The outcomes serve as focal points
for addressing the consequences associated with the use of the Benchmark assessment-curriculum innovation at both the Montrose and Arcadia research sites. They were selected as representative themes emerging from the data corpus in this study. This section also summarizes and concludes on the outcomes from the Montrose and Arcadia research sites.

i. Integration. The content knowledge, and the performance criteria used by teachers and students in the Montrose site were integrated prior to their use. As a result, when students were asked to "continue a story" for example, it was expected that they draw from knowledge across disciplines. Teachers from Montrose explained that this was the reason they chose monthly themes, e.g., Chinese New Year, around which their lessons and activities could be contextualized. These themes were general enough to allow for the integration of knowledge across disciplines, e.g., Language Arts, Science, Social Studies, Mathematics. As well, the Benchmarks (i.e., performance criteria) in each lesson were linked to the tasks and the content involved with the tasks. As a result, their seemed to be no demarcation between the assessment and instructional practices in Montrose. These observations help to confirm the seamless integration of content and performance standards in the Montrose site.

However, the same degree of integration between content and performance standards with the Benchmark innovation was not as evident in the Arcadia classrooms. What was being taught usually received the most emphasis; how it was done received much less emphasis. Students in these classrooms had few explicit criteria with which to apply their classroom performances. In the Arcadia classrooms, the assessment-curriculum Benchmark innovation was
more heavily weighted on the curriculum, i.e., on content, rather than on the assessment.

ii. **Alignment.** Alignment refers to the selection of teaching methods, classroom activities, and instructional materials appropriate to the students and the goals of the lesson. In the standards-driven Montrose classes, the lessons were objectives-oriented - students were explicitly aware of the purpose of each lesson, and were given examples of how the lesson tasks might be completed at various levels of performance. Students almost always worked in groups, and could choose from a number of activities associated with each lesson objective. They found some activities more suited to their interests and abilities than they did others. Classroom materials were not only plentiful, but also appropriate to many levels of student abilities. After students completed their Benchmark activities in these classes, they shared their work with their peers who discussed and critiqued the assignments. From these presentations, many students were left with ideas for improving their work. These ideas, along with the results of past performances, formed the basis for discussion in their twice-weekly student-teacher conferences. Students in the standards-driven classes saw the classroom instruction as tailored to their performance needs.

In the content-driven classes at Arcadia, the lesson focus was task-oriented - most of the instruction was directed to the tasks. However, the objectives for the lessons were treated implicitly, and students were not given examples of what constituted a well-done assignment. Many students found it necessary to interrupt the instruction and ask for more specific direction in their activities. As a result, some students were unclear as to how the classroom activities related to many of the lessons observed. In addition, students felt the teacher
instruction often wandered throughout the lessons - making the main purpose and topic unclear. Students usually completed their classroom activities in the content-driven classes without discussion with their peers or their teachers. Usually, all students were given the same activity to do - some performing better than others. In these classes, the curriculum materials were limited - forcing the teacher to use them with only a few students at a time. The school library offered little in the way of additional Benchmark resources for teachers' classroom use. When students completed their assignments, they would present their work to the rest of the class. However, these presentations did not include feedback from the other students in the class. Following their presentations, students would conference once per week with their classroom teacher. The purpose of these conferences was for the teacher to correct each student's work, and based on the results, assign remedial or enrichment exercises.

iii. Opportunity-to-Learn. Depending on the meaning one adopts for the term "opportunity-to-learn", marked differences exist in the way that instructional time was used in the standards-driven and content-driven classrooms in this study. If effectiveness is seen as a reference to how the program goals, objectives, activities and assessments are operationalized in the classroom setting, then the instructional time in the standards-driven classrooms were used very effectively in that the lesson objectives, activities and assessments were aligned with program goals. In the content-driven classrooms, such alignment was not as evident to the same extent. As a result, instructional time in these classrooms was not as focused on defined lesson objectives. Students, in fact, had a higher degree of ambiguity in what was expected of them in their assignments; they required much more clarification
of purpose after they had begun working on their assignments. However, if "opportunity-to-learn" is seen as a reference to "freedom to experience" the construction of a Language Arts project according to a prescribed format, then the time in the content-driven classrooms can be said to have been used effectively. Secondly, the division of student performance into a series of levels from novice to expert, each with identified dimensions, meant that all students in the standards-driven research site were continually engaged in the process of developing their levels of performance. None of them were seen as "failing to make the grade", but rather were identified for specific instructional supports designed to help them along with their performance. By comparison, the performances of students in the content-driven classrooms were not referenced to an identified range of development or expertise, but rather were assessed using a more limited and confined set of teacher standards. Consequently, if alignment between goals, objectives, activities, assessments, and student engagement are considered integral to instructional effectiveness, then students' time in the standards-driven classes was used more effectively than it was in the content-driven classes where alignment and a framework for student performance were not as developed.

Opportunity-to-learn also relates closely to instructional methodology and classroom resources that need to be combined with students' [engaged] time-on-task. Instructional methods in the standards-driven classes differed from those in the content-driven classes according to purpose, and frequency. For example, methodologies that were designed to integrate the Language Arts components of reading, writing, listening, and speaking in the standards-driven classes consisted of individualized and shared reading, storytelling and book talks, author studies, book talks and interpretive activities, literature dialogue journals, reading buddies, process writing, idea production and word
webbing, peer presentation and critiquing of each student's written work with subsequent revising, editing, and occasional publishing, and writing notebooks. These methodologies were framed within broad themes, e.g., Chinese New Year in which knowledge related to Language Arts could be integrated, and in which a students could choose from a wide range of activities. The activities for these students were open-ended, and consistently assigned so that students had daily opportunity to read, write, listen and speak. Students from the standards-driven classes almost always worked in groups, and could choose from a number of activities associated with each lesson objective. The materials available to students in the standards-driven classes were collected over a six-year period and included class sets of short story workbooks, literature, novels, factual books, and students' published materials. These materials were not only plentiful, but also appropriate to many levels of student abilities.

By comparison, classroom lectures accounted for the main instructional methodologies in the content-driven classes. Students in these classes were provided with an outline of generic questions that could be applied to any novel they chose to read and on which to report providing that it related to the chosen class theme, e.g., Olympics. Through lecture, teachers explained to students what was expected of them in regard to each of the "questions" on the novel study list. In the content-driven classes, the lesson focus was task-oriented - most of the instruction was directed to the tasks. However, because the objectives for the lessons were treated implicitly, and students were not given examples of what constituted a "well-done" assignment. Many students found it necessary to interrupt the instruction and ask for more specific direction with their activities. As a result, some students were unclear as to how the classroom activities related to the point of their lesson. Students
usually completed their classroom activities in the content-driven classes without discussion with their peers or their teachers. In these classes, the curriculum materials were limited - forcing the teacher to use them with only a few students at a time. The school library offered little in the way of additional Benchmark resources for teachers' classroom use. As a result, the students in the content-driven classes were expected to collect their own resource material for use in their assignments.

Further, students' opportunity-to-learn is linked to the evaluation strategies used to provide feedback on their classroom performances. Teachers in the standards-driven classes used a combination of holistic and analytic scoring devices with their students' work. Using established criteria for each task, teachers compared their students' performances based on students' past performances involving similar criteria, to their classmates, and to students across the Board. While the goals of the Language Arts program used by the Montrose teachers remained the same, the content and delivery of their programs were modified to address the range of students' performance levels.

Teachers in Arcadia found it difficult to shift from scoring students' work analytically to holistically. They found holistic scoring to be difficult. They weren't sure how to do it, were concerned about the time it might take, and were unclear about how they might use the assessment results for their instructional purposes. They limited the comparison of their students' classroom performances to other students within the same class. Consequently, their assessments had little influence on their instructional programs.

iv. Communication. Students from the standards-driven classes were aware of their Benchmark lesson objectives while doing the activities
associated with each particular lesson. Their standards for performance were articulated explicitly and shown to students on the overhead projector as a regular part of the introduction to each Benchmark lesson. The Language Arts tasks were expressed in terms of five levels of performance from novice to expert. Examples of how students' work at each level with the tasks in each Benchmark lesson were shown, and discussed with the students before they began working on their assignments. As a result, students in the standards-driven classes had a meaningful framework for judging their own work. They also knew what was required of them in order to reach each level of sophistication. Students viewed these levels as "starting points" for further improvement in the performance. In addition, the instructional objectives associated with each lesson were articulated to students as a regular part of the Benchmark lessons.

In the content-driven classes, standards for performance were stated in terms of content, but not in terms of performance standards. Students from these classes knew the "form" in which to deliver their completed task, but not the "function" of the task. For example, students gathered information on the Olympics, but were unaware of the characteristics associated with the levels of performance that ranged from novice to expert. In addition, the instructional objectives provided to students related to the "end product" rather than the process of being engaged in their tasks of completing their written projects.

One of the main dimensions of communication concerned the role of feedback in assisting performance, and in adjusting classroom activities as needed. When asked what information the Benchmark portfolio assessments provided regarding her students' performances and her classroom teaching, Muriel highlighted the importance of the sheet that summarized the performances of each of her students in detail:
...sometimes I think you can be fooled...tricked by thinking that the children are doing really well perhaps on being able to write the meaning of a story and using the same type of language and carrying on the same types of patterns that the original author used. But sometimes you just have to sit down and do a quick check system and find out that there might be three or four or five children hidden as a group at either one end that should be dealt with individually.

When asked if the feedback she received on her students' performance influenced what she teaches and how she teaches, her reply was:

Oh sure, absolutely. I think that you can look at and include all of those activities in your language program...but certainly the responses of the children...you can check the spelling to see if there are children who need to know the rule...you sometimes have to double the consonant before adding the ending - maybe it shows you that you need to that kind of a lesson, or maybe it shows you that you need to work on sequencing or maybe it shows you that somebody is ready to talk about similes or metaphors. It gives you all kinds of insights in to where you can go next, and it also shows you places where you haven't done as well as you thought you did or could.

When asked to describe the same question, Margaret from Montrose replied:

One of the things it does is point out to me which children can be grouped together for small-group instruction if I want to do them for ability or performance levels; another thing is it points out to me really clearly which kids are most at risk academically; it points out what kids perform quickly and easily which task you give them, and so those kids might need some enrichment or whatever; so it helps me to strategize what I'm going to do for each individual child as I go along. It also tells me if there's an area of curriculum that I might not have done a particularly good job on...so that would be an obvious outcome...For example, suppose that in doing something in the Chinese New Year, I ask the children to prepare a graph of some kind. I might say "Take all the books that we've got in our little [classroom] library on Chinese tales. Somehow sort
them out. Tell me your sorting rule. Then make up a simple picture-graph on the floor with your books". That would tell me right away whether kids can visualize a graph, whether they can understand the idea of a couple of axis on a graph, whether they have an idea of what the title of it should be or where you put numbers on a graph. And if I see very clearly that few or none of them can do that, then I know I need to do more work on graphing. So that would tell me weak areas on the curriculum.

When asked if the standards in the Benchmarks have provided a common reference point that gives understanding between the school and the students' parents, Muriel from Montrose replied:

Yes, it's definitely made my life a little easier. It took the pressure off of "Where is Johnny?" more to "Well where are you taking Johnny?", "What are you planning next?" and in actual fact, on the folder itself, all you have to do is lift your eye up one level and there are listed seven or eight skills to be worked on and developed to a higher level...so that is very straightforward, its right there and the parents know there is a program, and that there are standards, and that there are steps and what is coming next. Its focused on the positive, and it made it a report that dwelt more on progress and developmental learning than a lot of other interviews have done before.

Margaret explained how she involved parents in the process:

I show them their work, and I also show them the sample work in the Benchmark folder on the task done in class...and I say "Take a look at your child's sample, take a look at the samples in the folder, and see if you can match your child's performance with one of these in the folder"...and parents can readily see where their child fits. They can easily see where and why their child was given a level 4 performance rather than a level 5...they can see the strengths and weaknesses as readily as I can.

Margaret added that she can sit down with parents and explain that you do, in fact, have a set of standards. The parents can then benefit from knowing where their kids are at in terms of helping them with their work at home:
...you are looking at their child and comparing them with the standards that you do have in place, it seems to allay all kinds of fears that they seem to have about how they perceive the school system to be functioning...and they can say "Oh yes, you do have standards" and it seems to relieve them tremendously.

Muriel explained an important side effect of student involvement in their own evaluations in the Benchmark Program:

...its the perfect tool, especially with young children, to be able to sit down and talk about progress and look at samples from months ago and "Oh gosh why did I do that six months ago?" Parents...its the same thing, you know "Look at the difference, the change!"...and its good for teachers too. It also lets them know what the system standards are, and that there are standards out there and it allows all parties to compare the work against the standard.

From Arcadia, Annette stated that she had trouble applying information received from students' assignments:

I don't consider a lot of them meaningful as is...for example, L3-7, Listening, Narrative and Role Playing...some kids are extremely shy and don't want to have everyone else watching them. There's no way I'd do well on that...I don't think its fair. Level of comfort...some kids it takes a lot. It has nothing to do with how bright they are...you just can't hold it against them.

Teacher Adelle from Arcadia expressed somewhat more benefit to receiving information in connection with the feedback from the Benchmarks:

I guess in terms of where the kids are at grammatically and how the kids think. In re-telling the story...how they process the information down on paper...do they write in complete sentences, proper grammar. They help me know if I can integrate material".
When asked if she was personally satisfied with the usefulness of the information she received from the Benchmarks, Adelle replied:

Oh sure...because I think that if I'm having trouble with a child and the child has a lot of difficulty with writing in tone, or writing dialogue, or what re-telling is versus summarizing...I have something else to look at...because I do give these massive novel studies and when its a summary and question and answer, the child doesn't understand what he's put down on the summary because his question and answer tend to be really abstract or really something that very hard to look at in the summary. So you don't see the connection...the child needs to have a little more of a connection. So sure...you can go to the Benchmarks and say "I'm going to have him retell the story to see if he really understands what went on".

Students from Montrose. The students commented on the use of feedback for diagnosis and support as a regular part of their instructional program. The overall theme that emerged from the students in the standards-driven site regarding feedback was "how to improve on my work". Connected to this major theme, students expressed their views on the four dimensions of 1). self-assessment; 2). conferencing; 3). peer-editing; and 4). self-monitoring of their classroom performance. Student Michael from the standards-driven site expressed how he thought that the teacher's comments rather than the grade was most helpful:

I think its important to see your stuff to see where you can improve on it instead of just looking at your mark to see how good your mark is...because if you're just worried about your mark, you'll just keep on getting the same mark I guess because if you want to improve...then you'll keep getting better and better and your mark will get better and better as well.
Because students from Montrose were provided with the expectations and criteria for their performance, they were able to reference their marked assignments to these expectations and assess their work against these criteria. Mary explained that part of the feedback loop was the twice-weekly conferences that students from these classes had with their teacher:

We usually do conferences on writing...and our teacher reads this story to us and says "Okay, tell me what this is about"...she asks us some questions...like if you were going to Canada from Germany would you fly or would you take a boat or something?...and we would write about that later. So its a good place to get ideas for our work. And we also show our work...our teacher looks at it and sometimes says "Hmm...you might want to look up how this or that word is spelled"...or..."You might try to explain more clearly in your beginning what your story is really about"...So we go back to work on them...

Students from Montrose also read their written work to the rest of the class for feedback from their peers. Martin explained:

Sometimes my friends ask me questions about my story. And sometimes after I write a story in my book, you put it on the board and you can read it to the classroom...if you want to publish it you have to first read it to the classroom to see if its good and long. Like I had to read mine to the class...and I got it published. Sometimes they give you ideas if it needs to be changed...like my friend said to me...she said that "I didn't explain the conflict...what was the news"...I didn't explain the news!

The fourth dimension on feedback from the standards-driven classes concerned students' growing interest and ability in monitoring their own work. Mason commented:
Well, I got pretty much better in the second term than in the first term, and you can show that...you can see that because on the back of the [Benchmark] folder they have little squares and with your mark...how high...and I got a lot better in the second term.

Mary from Montrose offered another view to the importance of the portfolios in their class:

...it helps a lot because I know what I have to do to go up another level in my work. A lot of times I can just go ahead and work on it myself. And before it was harder to get our work, but now our teacher just puts our work in our folders, and we can go up and take a look how we did and work on assignments handed back.

**Students at Arcadia.** By comparison, the overall theme on feedback that emerged from students in two classes at Arcadia concerned the values attributed to assignment grades, and those attributed to teacher comments on assignments handed back to the students. Under this overall theme, students voiced the dimensions of 1). results obtained; 2). dialoguing for improvement; 3). keeping old assignments at home for personal assessment; 4). using classroom portfolios; and 5). participating in teacher conferencing. Student Andrew from Arcadia in the content-driven classroom commented that when he got his assignments back, he looked at his marks first:

Well, the first thing I do is go to the marks on the back, so I go and I want to see if its good...then I'll look through my scribbler because I want to see what mistakes I've made, then I would look at the comments she puts like: *Very well done!* or *Good sentence structure.* or *Good neatness, good organization.*

When asked if he would converse with his teacher about his assignment in order to improve Andrew responded:
Well, when I'm having trouble, I ask her how to do it, what we're supposed to do, how I can improve. And when...only when I'm having trouble I ask the teacher. Or if she's having a math lesson with the grade 5's...that's when I would ask somebody else that knows it...so if she's not available then, I would ask a friend that knows...

Allan from teacher Annette's class commented that feedback received from both marks and comments were helpful to him:

If she says it was creative, but it could be more organized, then that'll help me next time. Or if my writing was a bit sloppy, then the next time I would change that but if I keep getting As and I don't make any mistakes, then I'm not really going to learn much because people say "you learn from your mistakes"...

When student Alicia from Arcadia was asked if she knows how to improve her performance on classroom assignments, she replied:

Just by looking at what she said I'd probably know...and I'd go to her and ask her how I could improve Sometimes I also compare my work to my classmates to see how I'm doing.

Student Anton from Arcadia commented that with his old assignments:

What I do, at home, I have a file where I keep them all. I show my mom, and she and my dad they look them over and if I can improve, they tell me to do better in my mark. I just put it into my closet with all my other assignments that I have.

However, students interviewed from Adelle's class at Arcadia were somewhat more advanced in their use of the Benchmarks Program. When Amber was asked to comment on the feedback she received on her assignments, she replied:
Well, maybe because it tells me what I got and what I didn't. The most important for me is tell me where I need to improve. The mark...well for me, it just tells me how I've been doing...

When asked if she discussed the results of her assignments with her teacher:

Well, if you don't understand...then you go down on the carpet beside her and she gives us some examples...Let's say it was the word Valentines and she put the word on the blackboard and asks what we think goes with it...she started us off with "heart"...

As well, Amber commented how she thought that having her own folder of work [portfolio] in her class was beneficial:

I think its easier just to have one folder all to yourself instead of getting mixed up. And sometimes I look back at my work in my folder later on. I like looking back...lets say from the beginning of the year and see how I've been doing...how I'm writing differently. And sometimes when I need to write a story and I don't have any ideas...I go to my old stories to look at them to try and get ideas...

Another student (Ashley) from Adelle's class at Arcadia explains that when she conferences with her teacher each week:

If you have any difficulties...we talk to our teacher about them and she would explain how to do them...and when we finish the conference I take my story to the computer and finish it there before I read it...

Student Abdul explained other activities taken up in conferencing:

We have five stories to conference, and we have to pick one of those five stories to conference about...to tell her what I wrote about those stories...and why I wrote that. But now its getting a bit more tough because what we have to do is...she gives us assignments
for our writing and we have to put in citizens, villains and a climax and a male character, a female character, animal character...and you have to have an ending.

Regarding the concern of communicating performance levels to parents, Annette from Arcadia stated:

The issue is that the student might be in grade four, but is doing work at a grade six level. Even if the student's work needs more development at the grade six level, it's still advanced for that student! But, we have trouble communicating that to parents.

Adelle from Arcadia went on to comment more specifically some parents' concerns with this issue:

They make comments like: *They aren't exact enough...They're too subjective...too fuzzy...without any context.* For example, parents want to know, "How come my child only got a '2' on this assignment?" When I explain that the '2' refers to the level of work being done...it's actually very well done because the student is working on a Benchmark that is at an advanced grade level! But, most parents still think the '2' should be higher (at least '3') or its a failure and they think something is wrong! Parents want marks, parents want formal teaching, they want standardized testing...So, what I would like to see is that the presentation be a little clearer for teachers, and a few more criteria put in there that maybe teachers together can say: "This is why I see the level five".

v. Levels of thinking. Students in the standards-driven classes were encouraged to develop their thinking in conjunction with their levels of performance. It is within the design of the Benchmark innovation that content and performance standards were integrated with each other. First, students were encouraged to develop their skills and thoughts; the innovation was growth oriented. Second, as students progressed along the cognitive levels of
recall, interpretation, application, analysis and synthesis, and evaluation, they did so in relationship to the Language Arts tasks they were assigned. That is, the criteria for performing at the "expertise" level would involve, e.g., comprehension of text and main ideas, ability to use information, making inferences and drawing conclusions, organized writing, and sequencing of events and ideas. Therefore, as long as students progressed in the Benchmark Language Arts innovation, their thinking skills in reading, writing, listening, and speaking would develop as a matter of curriculum design.

In the content-driven classes, students did not have the benefit of a Language Arts curriculum that was designed to integrate content with performance standards. In addition, the levels of performance received from these students were left to variables out of the control of the classroom environment. For example, how students performed may well have been more a reflection of home than school support.

vi. Classroom roles. To begin with, teachers from the standards-driven site were providers of information. However, the length of class time devoted to this activity was considerably shortened in favor of students' new role of being active processors of the course content they received. Students complemented the core knowledge with their own in order to construct personal meanings with the tasks they were assigned. Teachers in these classes wanted students to process that which they learned. They viewed their instructional role as "facilitating" the students in processing knowledge. This view of instruction was consistent with students' conceptions of learning in the standards-driven classes as "understanding" and "application". Teachers from the standards-driven site also assumed the role of "teacher-as-planner" in that they helped to plan their curriculum, and organize their lesson activities.
Because the standards-driven teachers were closely connected to gauging students' classroom progress on lesson tasks, the role of "teacher-as-assessor" was also woven into their identity. Ultimately, these teachers also needed to assume the role of "resource developer" since they needed to collect and create a variety of classroom materials that students could use in completing their assigned tasks.

In the content-driven site, students were engaged in gathering knowledge for "presentation" in the form of a project assigned to them. Very little of the information was critiqued so that they could present it in a form tailored to their own levels of understanding. Teachers in these classes viewed their role as "providing" a knowledge base for their students. This view is consistent with their students' conceptions of learning in the content-driven classes as "to learn to do things". In this site, teacher activities only minimally touched on the roles of teacher-as-facilitator, teacher-as-assessor, teacher-as-planner, and teacher-as-resource developer.

vii. Concerns with implementation. Teachers from Montrose and Arcadia expressed common concerns regarding 1) the levels associated with the Benchmarks, and 2) the relationship between the performance levels and grade promotion. For example, Muriel from Montrose stated:

...in many cases, I think the standards are a bit low...and I don't know whether if its just that in this school children come from a more privileged background and they've had a lot of experiences or whether because the standards were normed on a big cross-section including ESL and special ed. kids...The kids I have, because they're virtually all English as a first language students perform very well for the most part.
Montrose's first administrator Mathews recounts that anything below a level five awarded for students' performances on Language Arts tasks was considered inadequate by parents:

When I left Montrose - we were still wrestling with that. What do you do with the numbers...because if it becomes A, B, C, D, E again and if parents think that every kid is going to get to level five...that's wrong. And what we were trying to do is get away from the numbers and talk about those levels...the content of those levels rather than the designation [number] and that's tough, but it's something you got to keep working on with parents. And one of the dangers is that parents think that level five, especially at a school like Montrose, if you're not at level five...something's wrong!

In turn, Annette from Arcadia commented:

I don't think the Benchmarks are very fair to students...some of them are pretty tough. I have four ESL students...they are usually bright but have trouble expressing themselves...For example, one student has good reading skills, but very poor writing skills. So, I'll need time to modify the Benchmarks to suit the reading level of some of these students.

Adelle from Arcadia expressed concerns with the level of standards set by the Benchmarks:

Well, I would like to be able to use the primary criteria for my primary kids, who are working at the primary level, and the junior level for the kids who are working at the junior level or some kind of mix and match. Now that might be the result of me not having had the training that everyone else has had...there might be some flexibility in that which I'm not aware of. But that's a frustration that I have where some of the stories that were chosen for this age level are far too difficult for my kids...for their listening and reading ability. But, having said that, since September, it's been getting gradually better.
Administrator Mathews from Montrose stated that the notion of "pass-fail" confounded the use of assessment standards in Montrose:

What's a pass?...Just saying levels one, two, three, four or five when teachers have done two Benchmarks, and if a youngster got a level two, and a level four, does it mean they've passed? There's still the old mentality that you've got to have seven out of ten to go from grade three to grade four...and that's what we have to work around. But the beauty of it is its one-on-one with the parent, looking specifically at that kids' work!

In addition, each participant in the Montrose and Arcadia research sites had their own specific concerns regarding the Benchmark innovation. Teacher Margaret and administrator Mathews in Montrose both voiced concerns with the areas of program evaluation, and varying perceptions of "growth" between teachers and parents. Muriel comments:

It's the program itself and how to do it, and how to use it, and how to evaluate it. That student voice in terms of the program needs to be opened up "How have you felt when you did these?" There's a lot of wisdom in what kids say.

Montrose's Language Arts Coordinator claimed that communicating students' performances to parents may be problematic due to differing perceptions of "growth". In turn, she suggested that the focus for learning can become deficit-oriented rather than growth-oriented:

As far as the teachers are concerned...its growth...it shows kids at levels on various tasks. In terms of the parents understanding, the teachers can say "He's here on this particular task, and that's fine, he'll catch up". That notion of "catch-up" is a developmental one, but whether parents are getting more hard-nosed about they can't add three and three or they can't spell anything except simple words, I'm afraid we're
going to spend a lot of time focussing on what kids can't do rather than what they can do...and that's a problem with evaluation...that we want to tell the parents what kids "can do", and the parents want to know what they "can't do"! So we need evaluation that can show you, find what the kids can do...and teaching has to be built on what kids can do and not what they can't do.

As well, a number of other concerns were voiced by Arcadia teachers and administrators. One of these concerns related to referencing the Benchmark standards to particular students. Adelle explained:

My major concern is that we can't generalize to the child...there are kids who need particular kinds of help, so you're always modifying it. And its time, I just don't have the time to sit down and modify it...read it with all the instructions.in a way that would really satisfy, and work to advantage for my students.

Another concern expressed by Alex was that of materials for the Benchmarks:

Yes materials...lack of enough materials. To do a whole class project is almost impossible because I don't have enough materials to go around for everybody. ...the supplies we have in our library...very, very small. I think there's about 50 books on the Benchmarks...for all levels, all teachers, and all students in the school. And I think the largest one [supply of any one book] is that one over there...The Whale's Song...five copies. For some, I have them doing silent reading and then doing something with that material.

According to Adelle, meaningfullness was an issue with the Benchmark materials:

It has to be meaningful. I think the best Benchmark I did was where the students had to retell their favorite chapter, from a novel...so some had done the vocabulary, some read it, some answered questions on it...so they knew it, so there was no anxiety there. They had to choose their favorite chapter, read it, and just retell that part. I thought that
was the best one I think and the kids really enjoyed doing that too. And it was something we were familiar with, so things went well.

When Arvel from Arcadia was asked what her experience with the Benchmarks was, she replied:

My concern with this is with learning disabled kids, I have to use the junior-school profile...because I've got a junior-school class...but I've got 12 year olds that are at a primary or pre-primary level...which is very frustrating. But by the same token, if the assignment says they must read it themselves and then retell the story, I have to read the story to them...

This teacher also expressed concern and difficulty with time needed for collecting and organizing materials to be used to support her Benchmark instruction:

I find the difficulty is picking..."How does this fit into my own program?" Maybe I'm doing it backwards, maybe I should be saying "I'll teach this first, and then give them practise at it and then test it".

Perhaps the most serious issue expressed by Adelle was the "how to" of the program, i.e., presenting it in their classrooms:

...how to present the material...I don't think it was presented properly to the staff...globally, not just in this school. So how to present this material without making it work in isolation...In other words, "Let's get some more staff development about how to make this less threatening, how to have this incorporated into the day-to-day program so nobody knows its a Benchmark. Because if you work the portfolio in isolation, it is...it's a writing folder and its a Benchmark folder and then we do our other work. Whereas instead, taking one of the folders from the back or that little project they did, and just put it in and just go through the project and say, "Well, I think this is L.13 - I think we're tending to just put the number down and say, "Oh well, I've done
that one, and I've done that one...instead of looking at the whole and saying, "Gee they've made notes, they've retold the story in written form it looks like, and they've written a letter to the editor of the newspaper talking about the novel...

Finally, all the teachers interviewed at Arcadia shared Annette's concern regarding the vagueness of the Benchmarks:

I think the whole concept is vague...the whole overview...why it was developed...who developed it ...why we feel it works...why we are accountable via this method...why its fair to all students...

viii. Reflections. When administrator Mathews from Montrose was asked for his reflections on developing the Benchmarks against the backdrop of previous programs he was involved with, he commented:

It was easier in the sense that you were working with something very specific and concrete and had an outcome...get it implemented. Whereas when you're dealing with moving from a traditional approach to teaching towards "How do we put into place what we know about kids into [teacher] practises ?" ...that's pretty nebulous. I mean that's in teachers' heads. So this was much cleaner. Looking back...we moved fairly well from developmental learning...to whole language teaching...to Benchmarks evaluation.

When asked if he looked back on the development with any disappointments, Mr. Mathews quickly responded:

One of the things I would do...in retrospect...is introduce it to the parents a little later. You can't have the parents ahead of the teachers...its not that they were, but I think we were saying things that weren't going to happen immediately. [Parents] wanted it to happen right then! And that's putting incredible pressure on teachers...I mean we were able to work around it but we realized after we had introduced it to parents much too soon!
Respectfully, Muriel, who contributed significantly to the Benchmark and Portfolio Programs, offered her reflection:

I think Benchmarks are a very unique method of evaluation...and I think its probably the best one I've ever been able to use because not only does it give you a chance to compare a child against a standard but it also gives you a chance to look at why they're doing what they're doing, how they're doing it...look at their written not just their reading skills...it allows you to look at language development totally.

Comments from the administrator and Language Arts Coordinator address the value of the program to the school, and the nature of its development in Arcadia. Arcadia administrator Ms. Ashton believed that the program is an improvement over other programs in the past. She stated:

When I think back to my teaching experience where the focus was on the Dolch basic lists. and you were still doing the standardized tests to see what level they were at...but it wasn't related to what they were doing in the classroom or reading in class. Even when you were writing up reports, it was important to have administered standardized tests to verify the teacher's evaluation of the child without the test. So what we have now in terms of assessment is something like "How the child performs on the basis of those things he does in class?"...not to say that this program is the best.

By comparison, the Board's Language Arts Coordinator was guardedly optimistic about the development of the Toronto Benchmark Program in Arcadia:

I would say that although it seems it may be rocky and slow and that nothing's happening, there seems to be so much misunderstanding, and you sometimes get so pessimistic about it, I think even so...given the pace at which most things get developed, its not that bad. I think now because of the portfolios, there's a concrete
piece of something that the teachers have to do. So even though they're uncomfortable with it, and maybe they're not going to do a great job in this round, at least they're trying it. With a lot of other things that people have tried to implement in the past, it never gets off the ground.

Arcadia classroom teacher Alex concluded:

I think this is one of the best programs to ever come out of the Toronto Board because it connects the objectives of instruction to the task to the students' assessments - a direct linkage. Its great just for teachers and for parents alone...but also for teachers and parents to communicate with each other on the students.

Summary. Students perceived the nature of instructional delivery in the standard-driven classes to be student-centred, i.e., shared by teachers and students as through classroom groups, teacher-student conferencing, and peer-edit sessions. Those in the content-driven classes perceived their delivery to be teacher-centred, i.e., with less opportunity for them to share in the instructional process, i.e., as through teacher lectures, and student seat work. At Montrose, teacher instruction conveyed clear and manageable lesson expectations to students. Classroom communication in the content-driven classes was more one-way, i.e., from teacher to students. In turn, students communicated individually in the content-driven classes with their teacher for clarification on their lesson expectations. In the content-driven classes, the lesson focus was task-oriented - most of the instruction was directed to "what" the tasks were rather than "why" students were to do them. The objectives for the lessons were treated implicitly. Many students found it necessary to interrupt the instruction and ask for more specific direction in their activities. As a result, some students were unclear as to how the classroom activities related to many of the lessons observed. In addition, students felt that teacher
instruction often wandered throughout the lessons - making the main purpose and topic unclear.

Students usually completed their classroom activities in Arcadia's content-driven classes by themselves. Usually, all students were given the same activity to do - some performing better than others. In these classes, the curriculum materials were limited - forcing the teacher to use them with only a few students at a time. The school library offered little in the way of additional Benchmark resources for teachers' classroom use. When students completed their assignments, they would present their work to the rest of the class. However, these presentations were usually not followed by discussion from the other students in the class. Following their presentations, students would conference once per week with the teacher during class. The purpose of these conferences was for the teacher to correct each student's work, and based on the results, assign remedial or enrichment exercises. In the standards-driven classes, the lessons were objectives-oriented - students were explicitly aware of the "why" of each lesson, and were given examples of "how" the tasks might be completed. Students almost always worked in groups, and could choose from a number of activities associated with each lesson objective. They found some activities more suited to their interests and abilities than they did others. Student materials were not only plentiful, but also appropriate to many levels of student abilities. After students completed their Benchmark activities in these classes, they shared their work with their peers who discussed and critiqued the assignments. From these presentations, many students were left with ideas for improving their work. These ideas formed the basis, along with the results of past performances, for discussion in the twice-weekly student-teacher conferences. Students in the standards-
driven classes saw the classroom instruction as tailored to their performance needs.

In the standards-driven classrooms, classes were routinely concluded with a summarization of the connection between the activities and purposes that transpired during each class. Such connective summarizations were comparatively rare in the content-driven classrooms. Students in each group had developed working and caring partnerships with their parents. Students communicated with their parents on their Language Arts classroom activities, and academic performances. In turn, parents served as tutors for students regarding their classroom performances. All students saw the support received from their parents as very helpful to their success in their Language Arts courses. In addition, students from both groups regularly borrowed books from their school and public libraries for home-reading and writing activities.

Included with each Benchmark used in the Montrose site were the criteria for five levels of holistic scoring together with a bar graph indicating the percentage of students who scored at each level. All the Benchmark criteria (Appendices J) were developed by extensive observation and examination of students' performances. These criteria reflect the intent of the objectives listed at the beginning of each Benchmark. The criteria also include many additional higher-level thinking, problem solving, and process skills. For example, Benchmark L3-6, levels five and four criteria include the skills of "making inferences", "drawing conclusions", and "organizing and summarizing information". In the Arcadia site, the lesson expectations were not often stated explicitly, and prior to the activities of each classroom lesson. As a result, the Toronto Benchmark Program standards stood in isolation, i.e., the purpose of students' classroom activities were not aligned with the
descriptive levels of students' performances on tasks related directly to the stated objectives of each Benchmark lesson. Arcadia students had no clear means of indicating the quality of their classroom performance as compared to others in the Toronto Board who completed the same lesson activities. In addition, students in the special needs classes were not provided much support with the Benchmarks as the tasks were developmentally inappropriate for them.

Teachers' perceptions of the assessment of knowledge gained prior to class revealed that the standards-driven classrooms identified a performance base for students. In turn, this "base" provided information that informed the instructional needs of students. By comparison, students in the content-driven classrooms entered into Benchmark activities without arriving at any information about the level of understanding and expertise of students regarding their currently-assigned tasks.
Chapter VIII

Analysis and Discussion of Findings

Data collected in this study have been organized according to research sites, antecedents, transactions and outcomes. The purpose of this section is to view the data in relation to the relationship between assessment and curriculum. Finally, the data from these categories are then related to the more general issues critical to this study.

1. Research Sites. Montrose and Arcadia were both designated as pilot schools for the implementation of the Toronto Benchmark Program. Both schools were within the Toronto Board of Education, house elementary students, and had approximately the same numbers of students and teachers. Differences between the two schools (as research sites) related to location and student characteristics. While Montrose was located in a very wealthy area of the Board, Arcadia was in an inner-city location. Students at Montrose have been referred to by teachers and administrators as being very motivated, and very engaged in their daily classroom and school activities. They were, overwhelmingly, uni-cultural with English as their first language. Occasionally students did transfer in and out of Montrose during the school year. Students from Arcadia, in contrast, were multi-cultural with English as their second language. Many of these students endured lengthy periods of travel to and from school as they lived in locations a fair distance from their school. Many of Arcadia's children lived with a single (and working) parent. These students had a high rate of transiency compared to those of Montrose.
Montrose and Arcadia were similar with regard to their professional concerns, and approaches to classroom performance. Administrators and teachers at both sites were concerned with the "looseness" of the evaluative part of student academic development. Participants from both schools were eager to make the instruction-learning-assessment triad more meaningful, and fluid. Finding an innovative assessment program that is student-centred, one that would complement the developmental and whole language approaches to learning and instruction was a goal shared by participants from each school. All participants thought locating students' levels of learning to be crucial to basing further growth in performance. As well, participants from both wanted instructional standards that could be used in comparing and communicating students' performances. Staff and administrators from both sites wanted to identify and develop the characteristics of good educational programs in order to promote their students' classroom performances.

However, the research sites differed according to the themes of administrative experience, documents used, and introduction to innovation. Montrose's two administrators provided a strong leadership role in developing the Benchmarks in their school. Montrose administrators were committed to the principles of developmental learning and whole language instruction. Particular events, from the 1960's on, in the professional experiences of both Montrose administrators, peaked their curiosities regarding student-centred education, e.g., staff communication, student learning, and assessment that would inform the educational process. The closely shared instructional beliefs and goals of both administrators allowed for ongoing development of the Benchmark Program in Montrose after administrator Mathews retired in 1991. The professional experience of administrator Ashton in Arcadia relevant to Benchmark development was limited to Board rather than school
participation, e.g., the Board Benchmark Development Committee. Arcadia administrator Ashton was in her second year of administration, all at Arcadia. There was no history of Benchmark interest or activity at Arcadia at the time Ms. Ashton assumed her position there. As a result, Benchmark development in Arcadia did not begin until 1992 (compared to 1989) at Montrose.

2. Antecedents: Precursors to Change. Table 8 helps to organize the participants' data which uncovers degrees of differences as well as similarities associated with the background and development of the Toronto Benchmark Program at the Montrose and Arcadia research sites.

Table 8.

Antecedents to Curriculum Change at Montrose and Arcadia.

<table>
<thead>
<tr>
<th>Antecedents</th>
<th>Montrose</th>
<th>Arcadia</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Conceptualization</td>
<td>Deep</td>
<td>Surface</td>
</tr>
<tr>
<td>2. Adoption</td>
<td>Firm</td>
<td>Tentative</td>
</tr>
<tr>
<td>3. Process</td>
<td>School-Based</td>
<td>Teacher-Based</td>
</tr>
<tr>
<td>4. Support</td>
<td>Widespread</td>
<td>Localized</td>
</tr>
<tr>
<td>5. Communication</td>
<td>Institutional</td>
<td>Individual</td>
</tr>
<tr>
<td>6. Staff Turnover</td>
<td>Discrete</td>
<td>Discrete</td>
</tr>
<tr>
<td>7. Teacher Resistance</td>
<td>Negligible</td>
<td>Overt</td>
</tr>
<tr>
<td>8. Portfolios</td>
<td>Instrumental</td>
<td>Rudimentary</td>
</tr>
<tr>
<td>9. Continuing Development</td>
<td>Provincial</td>
<td>Classroom</td>
</tr>
</tbody>
</table>
The characteristics of the themes distilled from the data helped to identify the similarities as well as differences with regard to the background and development of the Benchmark Program at each research site.

Participants' data from Montrose and Arcadia were most similar in the aspects of adoption, conceptualization, and staff turnover. Teachers from both research sites commented on the need to develop their own instructional program. They stressed the importance of keeping perspective in adopting any new innovation. They all believed that it was wise to look at any innovation cautiously, and take from it what can be of benefit, but still complementary to their own programs. The innovation, therefore, in this case was referred to as "added aids", and "program enhancers". No one interviewed [experienced or inexperienced] hinted even remotely that he/she would accept an innovation in its entirety, and simultaneously discard the program and teaching approach he/she had already developed. Nevertheless, when introduced to the Toronto Benchmark Program, participants in Montrose reported a very smooth handling of it: a "natural fit" with their ongoing whole language activities. The Benchmarks for them were a welcome addition: they were termed "the missing link" in the instruction-learning-assessment triad as it complemented developmental principles of learning.

Reports on the introduction of the Benchmarks from participants at Arcadia were not nearly so positive. References such as "...just threw it at us", "the program", "poor public relations", "lacking in rationale", and "they didn't even discuss it with us" are representative of staff views on the introduction of the Benchmark Program at this site. While the introduction of the innovation at Montrose was a "joint handling" of the Toronto Board and Montrose, staff at Arcadia received the program from the Board alone.
Montrose, at the time this data was being collected, was going to be losing seven teachers out of the 22 originally trained with the Benchmarks due to retirements, transfers, and occupational changes. The staff replacements would not have participated in the Benchmark development activities at Montrose. This has implications for classroom use and staff development. In Arcadia, occasional staff turnovers were not an issue as there wasn't the strong nucleus of teachers initially experienced with the Benchmark development that occurred in Montrose. However, in both cases, there was little Benchmark professional development provided for the new staff members.

The main differences regarding conceptualization of the Benchmarks related to teachers from Montrose identifying particular aspects of the Benchmarks, e.g., levels of learning, assessment for instruction, and their implications for classroom use. In Arcadia there existed no such degree of conceptualization of the Benchmarks, nor was there the focussed fit between the Benchmarks and portfolios. However, teachers from both research sites questioned the Board's top-down approach to introducing the Program, and the assumption that administrators would all buy in to the Program and help teachers put it into practice. All believed that more teamwork, at the classroom level, was necessary if better understanding of the program was to occur. Participants' reflections of the program in both sites highlighted the Benchmark Program as being an improvement over previous programs introduced by the Board; as having authentic norms for student performance; as being specific and concrete; as helping teachers learn more about the way in which students performed; and as aligning instruction, classroom activities, and student assessment.

Data from Montrose and Arcadia differed considerably more on six of the nine themes on development (process, support, teacher resistance,
communication, portfolios, and continuing development). The process of development in Montrose began with the "volunteering" of all staff in the school. Administrator Mathews began the year with development sessions every second week during lunch hours. He involved each staff member by having them chair the sessions. In Arcadia, staff were given much less direction, but more freedom in developing the program. In short, the Benchmark development in Montrose was mandatory while in Arcadia it was left up to the discretion of each teacher. As part of the development process, participants from Montrose utilized documents written by practitioners and researchers in the area of developmental learning (Observing Children), instruction (Language Across the Curriculum), and assessment (Measurement, Evaluation and Reporting to Children). These documents represented original sources of conversation and inquiry regarding Benchmark development amongst administrators and staff at Montrose. Only a few such documents were identified as sources for enquiry into Benchmark development at Arcadia. For example, Adelle at Arcadia used a booklet she created for herself (At a Glance) in which she recorded results of her students' performances while Alex used his document (Task Force Five) in order to conceptualize the change in teacher role from that of instructor to that of caregiver. However, these documents in Arcadia were not used as sources for discussion amongst the entire staff.

In Montrose everyone of the 22 staff members became involved in the process of developing the Benchmark innovation. During a two year "incubation period", each Montrose staff member did the same two Benchmarks in their class, and met regularly to discuss and improve on the associated instruction and materials development. Some teachers from Arcadia, in contrast, expressed difficulty in obtaining clarification on what
they were asked to do, what the Benchmarks were about, and why they were asked to develop it. More than any other variable, teacher resistance in Arcadia to developing the Benchmarks differentiated the two research sites from each other. For example, out of 27 teachers, only three completed more than the Board mandated two Benchmarks per year, and then discussed the results with their students' parents. When the Benchmarks were first introduced to Arcadia in 1992, a number of teachers refused to participate in its development. Most of these teachers did not practice whole language instruction.

Support for teachers in Montrose had been ongoing throughout the four-year span in developing the Benchmarks, and continues to this day. Examples of such support included Benchmark workshops for working through the lessons, grade-group meetings focussed for classroom use, and meetings with parents helped to explain the Benchmarks to them. In Arcadia, support meetings were not scheduled into the school day nor was any time provided on a regular basis for teachers to discuss and develop the innovation. With the shift in Board priorities from the development of the Benchmarks to the Common Curriculum, fewer professional days and resources were earmarked for Benchmark development.

Teachers in Montrose met and communicated regularly every two weeks to plan, share, and evaluate their progress in developing the Benchmarks. Currently (seven years after being introduced to the innovation, and one administrator later), they still meet once a month on scheduled time to communicate with each other on their use with the Benchmarks. No such meetings occur in Arcadia. Four junior-level teachers did try to meet on their own time to communicate their progress with each other, but those meetings gave way to other scheduled duties. Communication on the Benchmarks in
Arcadia is now based on impromptu discussions amongst teachers based on their informal social networks with each other.

The Board-accepted portfolio was developed by a teacher in Montrose to address the issue of recording students' performance results and their communication to parents. These portfolios were accepted for use by staff at Montrose early in their development of the Benchmarks. Some teachers in Arcadia used bins for storing student work, booklets for recording marks, but none of the teachers interviewed and observed there have integrated the portfolios with their programs of instruction.

Continuing development of the Benchmarks in Montrose has expanded into other academic subjects - particularly science, mathematics, and also French. As well, teachers from Montrose regularly provide inservices and workshops to teachers in neighboring schools who are not as far along in developing the Benchmarks. The thrust for continued development in Arcadia has centred on integrating the Benchmarks with the new Common Curriculum; in effect, the development of the Benchmarks was being de-emphasized in that site.

3. Transactions: Assessment and Curriculum in Operation. The major themes gleaned from the classroom observations of the Montrose and Arcadia research sites centred on 1). the expectations for student performance; 2). the relation of classroom activities to lesson objectives and assessments; 3). the delivery of the Benchmark lessons in class; 4). the integration of standards concerning course content and student performance; and 5). the manner in which student progress was monitored. Table 9 identifies the dimensions of these themes at each research site.
Table 9.

Transactions: Characteristics of Lessons in Operation

<table>
<thead>
<tr>
<th>Transactions</th>
<th>Montrose</th>
<th>Arcadia</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Expectations</td>
<td>Explicit</td>
<td>Implicit</td>
</tr>
<tr>
<td>2. Activities</td>
<td>Aligned</td>
<td>Irrelevant</td>
</tr>
<tr>
<td>3. Delivery</td>
<td>High Application</td>
<td>Limited</td>
</tr>
<tr>
<td>4. Integration</td>
<td>Content &amp; Performance</td>
<td>Content</td>
</tr>
<tr>
<td>5. Progress</td>
<td>Ongoing</td>
<td>Discrete</td>
</tr>
</tbody>
</table>

In the Montrose classrooms, the students' expectations for performance were explicitly identified and discussed on two levels. First, students in Muriel's class in Montrose read aloud the first three pages of the story *Chin Chiang and the Dragon's Dance* in which a problem was posed for Chin Chiang. Students were asked to complete the narrative and express in writing how Chin Chiang might solve his problem. Second, Muriel's portrayal on the overhead projector of three examples of performance in completing this narrative helped her students become more aware of the criteria associated with good writing. By comparison, none of the four teachers at Arcadia used the overhead projector [or any other method] to identify specifically what students were expected to do in each Benchmark lesson observed. Students in these classes were given, at best, general expectations for performance, e.g., "create a crossword puzzle based on a book", "design a book jacket" or "create an activity of your own". In addition, these general expectations were given without the criteria underlying the performance and assessment of their work.
As part of their classroom activities, Montrose students related the text of the story to their own personal experiences, and extended their ideas so as to inform, to explain, to describe, and to give voice to imagination and fantasy. At the same time, students demonstrated patterns of spelling, appropriate grammar, and rules of punctuation required for clarity of expression. At the end of the lesson, students were assessed according to the Benchmark criteria concerning their understanding of the characters and events in the story, their awareness of how their endings might be best received by an audience, and how well their writing was expressed in terms of organization, style, and rules of grammar. The classroom activities of students in the Arcadia classrooms observed were not tied as tightly to the objectives of their classroom lessons. As a result, there appeared to be a much looser fit between purpose and activity in these classrooms. To me, the question that seemed to unfold was: *How far along the line of explicitness must a teacher go in linking the objective(s) and activity(s) of each lesson?* As it was, students in the Arcadia classes were very engaged in their classroom work.

Concerning the delivery of the Benchmark lessons at Montrose, the expectations for performance in lesson activities were administered without any problems experienced by the students. This pointed to the relative homogeneity and high academic abilities of the Montrose student body. In the classes observed in Montrose, no special needs students with academic abilities less developed than his/her peers were present. In fact, the desire amongst almost all students in these classrooms was to reach expert levels of performance on their Benchmark activities. To this end, a number of these students had families who paid for out-of-school tutors to help develop the students' skills in reading, writing, listening, and speaking. In the classes observed at Arcadia, teachers were very limited in taking, as is, the published
Benchmark lessons and applying them to their classrooms. Clearly, the Arcadia classrooms contained a much more socio-economic, and ability diversity amongst its students within and between classrooms. It was more the rule than exception that teachers in Arcadia expressed dissatisfaction with the appropriateness of the Benchmark lessons. For example, even though Arvel had students who were twelve years of age, they were learning-disabled and had not yet developed their reading skills to a point that would enable them to complete the "silent reading" Benchmark lessons: in fact, their teacher needed to read the story(s) to them. In other cases, Alex was concerned about the lack of availability of curriculum resource materials critical to teaching the Benchmark lessons to her students while Adelle simply labelled the Benchmarks as "unfair" in that they were out-of-synch with her students' English language development, and lower socio-economic class. In short, teachers at Arcadia were forced to modify the Benchmark lessons and develop a set of performance criteria more in keeping with the characteristics of their students.

In Montrose, the classroom transactions were based on the integration of explicit content and performance standards in the innovative Language Arts Benchmark Program. Any teacher using the Benchmark innovation as intended would, as a matter of design, incorporate content with design [or vice versa] in his/her instruction and assessment practices. The curriculum at Montrose was standards-driven: the reading, writing, listening, and speaking tasks assigned to students were linked to specific criteria for performance. In effect, the criteria [i.e., Benchmarks] associated with these tasks differentiated students' performances on a continuum from novice to expert. As a matter of course, students and teachers alike could relate classroom performances to where they were located on the continuum. At this juncture, two educational
purposes were served: 1). students' performances were assessed through a framework containing criteria at various levels of expertise; and 2). information was given to the classroom teacher regarding the form and function of instruction appropriate to the continued development of the students. The critical components of this entire process were the Benchmark criteria made available to teachers and students, and the feedback available to them. At Arcadia, there was very little integration of content and performance standards: 1). students performances could not be assessed by as comprehensive a framework of criteria that related to their assigned tasks; and 2). teachers were not provided with the same kind of instructional information as were their colleagues from Montrose. Clearly, while the design of the Benchmark lessons promoted the integration of content and performance standards, the actual integration of these two sets of standards would occur only if the Benchmark lessons were adopted in their entirety.

The last theme that evolved from the observations of the classroom sessions concerned the way in which student progress was monitored. At Montrose, for instance, the Benchmark lesson Chin Chiang and the Dragon's Dance was used once before with the class observed, and would be used again with this class later on in the school year. Even though the same lesson was used three times, the Benchmark criteria for student performance related to this lesson would still apply: what differed on each occasion was the curriculum materials that would help students fulfill the objectives of this lesson. As a result, the teacher of these students was able to identify, remediate, and report on her students' performance with regard to the same Language Arts criteria on an ongoing basis. Because teachers at Arcadia used course content more as discrete bits of knowledge with their lessons, the information collected throughout the year that reflected the performance of their students was more
related to a collection of results on single trials of classroom activities loosely connected to each other. The main difference was that the assessment data of the students from Arcadia did not reflect a more continuous development of Language Arts skills based on the performance criteria that remained constant.

4. Outcomes: Measure of Performance. For the most part, both of the Montrose and Arcadia research sites' outcomes centred on the clarity of the Benchmark standards, and their specific relationship to grade-level promotion. However, the outcome concerns expressed in Table 10 also reflect differences in the stages of development for each site regarding their involvement with the Toronto Benchmark Program.

Table 10.
Assessment -Curriculum Outcomes.

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Montrose</th>
<th>Arcadia</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Opportunity-to-Learn</td>
<td>enhanced</td>
<td>partial</td>
</tr>
<tr>
<td>2. Communication</td>
<td>feedforward</td>
<td>feedback</td>
</tr>
<tr>
<td>3. Levels of Thinking</td>
<td>procedural</td>
<td>declarative</td>
</tr>
<tr>
<td>4. Classroom Roles</td>
<td>multiple</td>
<td>single</td>
</tr>
<tr>
<td>5. Concerns with Use</td>
<td>refinement</td>
<td>adoption</td>
</tr>
<tr>
<td>6. Reflections</td>
<td>certainty</td>
<td>vagueness</td>
</tr>
</tbody>
</table>

Of all the outcomes identified, those of 1). opportunity-to-learn; 2). communication, 3). concerns with the implementation, and 4) reflections were addressed by the participants most frequently, and most in-depth.
Together, these four outcomes served clearly to differentiate instruction and assessment practices in the classrooms researched.

The opportunity-to-learn outcome was a main theme amongst participants' responses, and researcher classroom observations as well. To provide a mechanism for collecting information which would reveal the effect of instruction on student performance, researchers (Carroll, 1963; Keeves, 1976) introduced an opportunity-to-learn (OTL) concept. Defined as "time actually spent on the act of learning" (Carroll, 1963, p. 725), OTL basically holds that, other things being equal, the degree of learning is strongly a function of the amount of time which pupils actually engage in learning. Consequently, it may be assumed that with more time on learning provided, learning would increase. More recently, researchers (Porter, 1994; Smith & O'Day, 1993) have tried to expand the concept of OTL to include the variables of educational standards, and school resources. In the two research sites in this study, students had access to classroom materials, and had time to engage in their classroom performances. The difference was the degree of focus on the learning tasks, and the focus on the supports for students to accomplish their tasks. In the Montrose site, students were given not only specific objectives, but also examples of the range of expertise in fulfilling their objectives. Students in Arcadia, though provided with lesson objectives, were not given examples of the range of performances associated with meeting their objectives. In short, while content and performance standards were integrated at Montrose, students in Arcadia received instruction emphasizing content.

Participants reported that performance expectations were delivered more clearly to the Montrose than to the Arcadia students. While students from Arcadia received feedback (Ramaprasad, 1983) from their assigned tasks, those from Montrose received feedback and 'feedforward' (Torrance, 1995),
i.e., the feedback was incorporated into the daily lessons in order to enhance students' learning. The information given to Montrose parents on their siblings' classroom performances was received more clearly, and more meaningfully than that received by Arcadia parents. A related difference between the two research sites existed in terms of the use of their classroom communications. In the Montrose site, teacher assessments were designed to yield evidence on student performance that were used to interpret how well students were meeting their lesson objectives, and how they could be supported in *continuing to improve* to meet these same lesson objectives. In contrast, the teacher assessments at the Arcadia research site were treated more as "ends-in-themselves" rather than as "means-to-an-end". Consequently, the focus on student development was more of an emphasis at Montrose than it was at Arcadia.

The third area of marked difference in the outcomes listed concerned issues with the implementation of the innovative Language Arts Program itself. Clearly, teachers from the content-driven research site were not as conceptually aware of the purpose, and the operation of the innovation as were those from the standards-driven site. Closely tied to this difference in conceptualization were corresponding differences amongst the content-driven participants related to time to prepare for the use of the innovation, professional supports available, and willingness to adopt or even adapt the innovation for use in their classrooms. Data from the participants also revealed that the content and performance standards were more integrated in the standards-driven than in the content-driven research site; that the intended, the taught, and the tested curriculum were more aligned in the standards-driven rather than the content-driven research site; and that students in the standards-driven site were encouraged to perform at more diverse levels of
thought than were their counterparts in the content-driven site. Teachers in the standards-driven site assumed a greater variety of educational roles for their students in addition to that of "delivering" knowledge than did their counterparts in Arcadia. The standards-driven participants' reflections of their experiences with the assessment-curriculum innovation highlighted the relative ease with progressing from a developmental learning philosophy, to whole language teaching, to Benchmarks use, to portfolio assessment. Reflections of teachers from the content-driven site dwell on the high number of curriculum initiatives to be dealt with simultaneously, and the high level of misunderstanding regarding the purpose and use of the Language Arts innovation.

The reflections of administrators and teachers also represent major differences in the implementation of the Benchmark innovation at both sites. Participants at Montrose saw the innovation as something very concrete rather than nebulous, i.e., they had a good understanding of how teacher instruction and student assessment were linked to achieving the objectives stated in each Benchmark lesson plan. Although comments by participants at Arcadia regarding the Benchmark program reflect their beliefs about the positive value of the Benchmark innovation, most of their reflections speak to the misunderstandings with the innovation itself, and the difficulties experienced in implementing it.

The remaining two themes of levels of thinking, and classroom roles complete the comparison of outcomes in the two research sites. There existed more emphasis at Montrose regarding procedural or higher-level (Bloom, 1956) thinking than at Arcadia. As a function of the instruction and assessment experienced at Montrose, students were required to perform at levels of making inferences, application of the information they were using,
and making judgements on that information. In Arcadia, students spent more time in recalling, and interpreting the Language Arts material they collected.

Teachers' educational roles in the two research sites differed according to purposes of instruction. In the Montrose site, students were expected to process the knowledge associated with their Language Arts curriculum. Therefore, teachers at Montrose acted not only as providers of knowledge for their students, but also as facilitators of learning for their students. Consequently, teachers at Montrose acted as planners, assessors, and resource developers. Because the main purpose for instruction and assessment was the retention of knowledge itself, teachers in Arcadia acted mostly as providers of knowledge, and minimally as facilitators, assessors, planners, and resource supports.

It is important to stress that the above antecedents, transactions, and outcomes are closely related to the contexts of the Montrose standards-driven and Arcadia content-driven research sites themselves. In the standards-driven site, teachers and administrators 'incubated' and 'developed' the Language Arts innovation for two years prior to introducing them into their classrooms for use with their students whereas in the content-driven site, teachers were expected to begin using the innovation almost immediately after being introduced to them by their Board. At the time of data collection for this inquiry, teachers in the Montrose site had been using the innovation with their students for two years whereas teachers in the Arcadia site were also into their second year of use. However, the classroom use of the innovation was much more sporadic in the Arcadia as opposed to the Montrose site: the administration in the Montrose site had encouraged and required teachers to use the innovation while the administration in the Arcadia site had not placed teachers under the same requirement. In addition to financial, the Montrose
site also had access to more human resources than did the content-driven site in implementing the Language Arts innovation. Accordingly, use of the innovation in the Montrose research site was associated with personal and institutional commitment, along with time and resources for development. In the Arcadia site, the period of development for use was extremely short, and uneventful; and implementation was characterized by sporadic adoption, use, and understanding.

As well, it is important to underline that the Montrose and Arcadia classrooms expose differences in purposes and procedures related to educational values. In the Montrose classrooms, the emphasis placed on learning meant that the purpose of instruction was for helping students become better processors of knowledge through classroom activities that challenged them to construct their own meanings of knowledge. In contrast, the Arcadia classrooms emphasized the value of knowledge for its own sake. Consequently, the purpose of classroom presentations was for students to "receive" the knowledge imparted and to reinforce this reception through classroom activities that encouraged students to remember the knowledge given to them. Also, student activities in both the Montrose and Arcadia classrooms were generally open-ended; i.e., students had opportunities to construct their own meanings to task situations. The difference was that the activities in Montrose were referenced to explicit levels of performance and were based on specific performance criteria that related to the stated objectives for each Language Arts Benchmark lesson. The Arcadia lessons did not have the same specificity and clarity, for student performance, and for the assessment of their performances.
5. Critical Issues. This study has shown that it is possible for assessment practices to be associated simultaneously with the goal of measurement, and the goal of instruction. Teachers in Montrose who had the time and resources to develop, understand, adopt, and use the Benchmark innovation were able to combine the functions of formative and summative assessment in their classes. For example, Muriel in Montrose was able to gather evidence on the performance of her students that enabled her to judge the extent of her students' performance in Language Arts, as well as being able to redirect this evidence for purposes of advancing the performance level of her students.

However, the issue underlying the relationship between formative and summative assessment needs to be clarified. Researchers (Black & Wiliam, 1996) locate each form of assessment at opposite ends of a continuum. At the formative end, the problem of creating shared meanings beyond the immediate setting are ignored: assessments are evaluated according to the extent they provide a basis for successful action. At the summative end, shared meanings are much more important, and the distortions and often undesirable consequences that arise are usually justified by appeal to the need to create consistency of interpretation. In short, when formative functions are paramount, meanings are validated by their consequences, and when summative functions are paramount, consequences are validated by meanings. The tension created rests on the perceived purpose(s) of instruction and assessment: that of learning, and that of measurement.

A number of suggestions might be presented to ameliorate this tension. Based on the claim that the formative and summative functions of assessment require different approaches to data collection, and that each approach will color the interpretation of the data, one might suggest that the gathering of data for student assessment be separated from the interpretation of data. It is
not too difficult to see that in deciding how the two might be separated may vary directly with the interests of the participants involved in the decision-making! For example, in some instances, teachers might be limited to collecting, but not interpreting the data generated from classroom testing. Although research (Brown, McCallum, Taggart, Branson & Gipps, 1995) has shown that this option may not be supported by teachers, in practice it is being implemented. For example, in the Province of Ontario, the Education Quality and Accountability Office (EQAO) has administered province-wide testing for 147,000 grade 3 Ontario students in Language Arts and Mathematics. In keeping with the research by Brown et al., many of the teachers (e.g., 93% of the 1124 grade 3 teachers surveyed by the Ontario English Catholic Teachers' Association) whose students were tested by the EQAO mounted public criticism of such large-scale external testing on the grounds that 1) there existed a poor curricular validity between the items in the test, and the content covered in the classrooms; and 2) there were existed concern regarding the use of the data collected (L. Amato, personal communication, Feb. 20, 1998). These concerns are not without precedent in our nation's schools. The publication of "league tables" (published lists of students' results on ministry examinations that rank school boards and schools based on students' scores) in Quebec has created its share of controversy (Bacon, 1995). Maheu (1995) reports that socioeconomic context was not accounted for with the publication of Quebec's "league tables" - a factor believed to contribute to the potential for the misuse and misapplication of examination results. Such publications added to the concern about test misuse and misinterpretation, and also contributed to the development of the Principles for Fair Student Assessment Practices for Education in Canada (Rogers, 1993). These principles state
specifically that developers and users of test data must address factors that might have had an influence on the test results.

Black (1996) expresses some doubt that the measurement and learning functions of assessment can be reconciled in a single curriculum. He goes on to suggest that perhaps the only way to preserve learning will be to lower the stakes, to trust schools and teachers more wholeheartedly, and to stop demanding from the school system an accountability product which cannot be manufactured without damaging that which it is meant to improve. However, this study has shown that it is, indeed, possible to combine the formative and summative functions of assessment simultaneously. Muriel and Margaret in the Montrose site were able not only to assess their students according to performance criteria specified as critical to the Grade 3 Language Arts program (see Appendix J), but were also able to measure the performance of their students against a backdrop of performance results (Benchmarks) compiled from students across the Toronto Board: Benchmarks which were expressed as levels of achievement ranging from novice to expert (see Chapter VI in this study, The Shift to Performance Assessment). It is this combination of content and performance standards that have been integrated into a program in its design stage of development that represents a possibility for combining the instructional and measurement functions of assessment.

Once the competing goals of assessment are addressed, the three remaining issues of assessment identified at the beginning of this study become more readily addressed. For example, the particular emphasis placed on the curriculum, and on student assessment becomes a moot point: they assume a role complementary to each other so that through the interaction of the two, they provide mutual support to the development of students' classroom performance. To administrator Mathews at Montrose, there were no rough
edges in this regard. To him, the relationship between the curriculum and the assessment was seamless: it was an instructional program just as much as it was an evaluative program. In addition, comparing student performance at Montrose was also axiomatic, students' results were compared to their own [previous] results on the same Language Arts tasks, to those of their class, and to those of students in the same grade across the entire Toronto Board. Such information provided teachers with the framework for communicating students' performance results to the students themselves, their parents, their school administrator, and to other schools within the Toronto Board.

Most of the emphasis in the Arcadia classrooms with the Benchmark innovation was on curriculum content; students' classroom performances were compared to each other in the class; and students' results were communicated to parents based on the arithmetic mean of students' discrete performance results over time. This was not a "black-and-white" comparison between Montrose and Arcadia: in fact, Annette and Adelle from Arcadia were beginning to develop a dynamic relationship between their instruction and their assessments of students. They were beginning to compare student performances over time in relation to stable performance criteria identified in the Benchmark Program, and they were finding their communications with parents and the school administrators regarding their students' progress more meaningful. The difference was one of degree, but this difference of degree redirects the attention from issues of assessment to issues of educational change.

The issues of educational change provide the most striking differences in the contexts, antecedents, transactions and outcomes associated with the assessment-curriculum relationship at the Montrose and Arcadia research sites. Of the five issues outlined in the beginning of this study as critical to
educational change, one has emerged as pivotal to the others: the process of change. When the Benchmark innovation was undertaken in the Montrose research site, it represented another link in the chain of developments already in progress at Montrose. That is, Montrose administrator Mathews and his teaching staff were already focussed on the idea of learning as a developmental process, and on providing whole language instruction for students. On the basis of an instructional approach that integrated knowledge across subject disciplines, the Toronto Benchmark Program extended the impetus in Montrose towards integrating content with performance standards. Soon after, the Montrose administration and teaching staff went past the Benchmark innovation by developing and using the Benchmark portfolio system to provide the missing link in the assessment-curriculum relationship: that of connecting teacher instruction, and students' classroom performances with communication to the students' parents, and to other schools where the students might attend. While the innovation was "on line" with the path travelled on by the administration and staff at Montrose, the same "on line" events did not precede the introduction of the Benchmark innovation at Arcadia. Teachers at Arcadia were used to developing their personal styles of instruction in the midst of multiple Board mandated innovations. Briscoe (1993) suggested that whether a teacher is able to adopt and implement alternative assessment routines as part of their instructional approach consistent with goals for meaningful learning is strongly influenced by what the teacher already knows or understands about teaching, learning, and schooling. Personally constructed beliefs, e.g., assessment as a system for reward or punishment, emerged in this study as part of the knowledge teachers used to test the fit of new ideas or possible actions as they attempted to develop and use alternative assessment practices.
Hargreaves (in Hargreaves & Fullan, 1992) reinforces this finding:

Teachers teach in the way they do not because they have or have not learned. The ways they teach are also grounded in their backgrounds, their biographies, in the kinds of teachers they have become (p. ix).

In addition, research by Torrance (1995a) is consistent with the data from this study that reflects teachers' high values placed on their instructional styles. Torrance noted that the expectation of influencing teaching and learning is often foremost in arguments for new assessments. His review of evidence based on teacher involvement with the General Certificate of Secondary Education (GCSE) in England and Wales suggests, however, that changes in assessment are unlikely to have the expected impact if teachers are not engaged in ways that help them see the assessments from a teaching perspective rather than only a testing perspective. He suggests that there is a need to go beyond simplistic notions of measurement-driven instruction and to think of authentic assessment "as providing a new framework for the discussion and development of instruction" (p. 56). In Arcadia the Benchmarks were seen as another Board mandated innovation, independent of teachers' approaches to instruction, that would surely pass by as did previous innovations. As a result, the process of change was seen to be "top-down" at Arcadia, with few resources allocated to developing the innovation. At Montrose, the difference was that administrator Mathews first made a [unilateral] decision to proceed with the Benchmark innovation, but immediately after this decision involved his teaching staff in the development of the program at his school. In effect, the change process at Montrose was highly school based. Conceptual change research (Posner, Strike, Hewson & Gertzog, 1982) suggests that lasting
change involves teachers reconciling their beliefs with the information gained from the implementation experience, and being able to reduce inconsistencies with existing beliefs by constructing alternative beliefs that seem meaningful and functional to using the innovation. Further, Tobin and Jakubowski (1990) assert that teachers must make a conscious effort to reflect on inconsistencies among the beliefs that are used to make sense of the assessment practices. As a result, it only stands to reason that if teachers are expected to change their assessment practices, those who assist in that change must provide opportunities for teachers not only to make explicit their personal beliefs, but also to reflect on them. In supporting, teachers' experiences that assist in the reconstruction process, it also stands to reason that encouragement and resources be made available to continue the innovation after it has been implemented in order to promote changes in teachers' attitudes and perceptions. Guskey (1989, p. 445) argues strongly that "significant changes in teachers' attitudes and perceptions is likely to take place only after changes in student learning outcomes are evident". As well, Mayfield and Weston (1996) studied thirteen third grade classrooms where teachers were using externally-mandated performance assessments as part of their instruction. Although one group of teachers received professional development sessions to help them along with their classroom use of the new form of assessments, the researchers found no significant difference in use between participating and control schools. Although there was some evidence of change made by the spring term, several teachers reported that they did not fully understand and adopt project ideas and assessment strategies until they began planning and thinking about what and how to teach in the following year. This view is consistent with the findings in this study: fundamental and conceptual change occurred slowly. Further, changes in student performances must necessarily
come last, after changes in teacher thinking and changes in instruction. In other words, only after teachers realize the innovation is plausible and fruitful can change in their attitudes, perceptions, and/or pedagogical mental models be expected.

The nature and extent to which the teachers in each location were able to develop the innovation; the roles they enacted when teaching it; the control they had over the form of the innovation; and their conceptualization of the innovation all served to influence their adoption and use of it in their classroom settings. Tables 1 through 4 (Fullan, 1982; Smith & Tyler, 1942; Fullan & Pomfret, 1975; Berman et al., 1976) in the literature section of this study identify and clarify the importance of such variables as leadership, beliefs, time, information, participation, feedback, and communication to the conceptualization, adoption, use and institutionalization of a curriculum innovation. Further, Tables 8, 9 and 10 organized under antecedents, transactions, and outcomes in this chapter provide specificity to the nature of experiences for Montrose and Arcadia participants with curriculum and assessment.

It seems that the literature on school-based curriculum development (SBCD) is particularly relevant to which consequences on curriculum and assessment might emerge as a result of the antecedents, transactions, and outcomes associated with a curriculum innovation. SBCD (Brady, 1990) relates directly to a school's capacity to plan, organize, lead, and evaluate curriculum change. Although capacities that exist between schools to manage curriculum change may vary according to their visions, intentions, and resources available to enact that change, SBCD is particularly relevant to this study in that one school (Montrose) demonstrated the progress that could be made when teachers are involved in the decision-making process related to
curriculum development and implementation. In addition, the participants in the Montrose site also demonstrated that they were able to form a partnership (Fullan, 1992) with their Board in proceeding with the Benchmark innovation. As a result, Montrose was able to gain access to much of the Board's resources in developing, and implementing the Benchmark innovation. By comparison, the in-school development of the Benchmark innovation at Arcadia was limited to only a few teachers, and with very little partnership with the wider Board in which it was located.

Analysis of data in this study accentuates the notion that the merging of the formative and summative functions of assessment, and the process of curriculum change is not a simple mechanical process. It points out that participants involved in the change process need to be concerned with issues of process and assessment alongside issues of content. In addition, it points out that within the design for the curriculum we have to acknowledge our intentions for change. Perhaps most significantly, the data in this study has pointed out that change is not simply a technical problem, but a cultural problem that requires incorporation of the contexts and shared meanings of the participants involved in the design of the change planning process.
Chapter IX

Conclusions, Implications, and Recommendations

The conclusions in this study relate to the "learning what" and the "learning how" of educational change (Nias, 1993). The former concerns the knowledge of the assessment innovation while the latter concerns the procedures to be implemented in order to operationalize the innovation. However, it is the "learning how" that is emphasized in this chapter since the findings concerning the assessment practices are consistent with those from other studies, and those concerning the implementation process offer new glimpses into the management of educational change.

With the shift away from traditional testing, as in the Toronto Board, to more contextual-based performance assessments, the Montrose participants confirmed that the expectations for success are made explicit according to observable exemplars and descriptions of students' work. What is helpful, from a teacher's and a parent's view, is that these exemplars become the standards. For the most part, such standards were seen by teachers, students, and parents alike at Montrose to be credible as levels of performance (in contrast to arbitrary standards set by textbook publishers, parents, teachers or school administrators) which can be used to ascribe the level of performance to subsequent performances on like tasks. No matter where a student's performance falls in relation to the standard, performance is measured by determining how students perform in relation to the descriptive standards over successive trials. However, concerns for reliability and validity of the standards need to be met. Although the literature speaks to this concern, more
attention must be directed to improving our ability in setting standards, and in measuring how they are achieved.

Participation in the development of curriculum offers teachers a role in shaping educational programs, a sense of involvement and responsibility in the implementation process, and a commitment to the success of the program (Morin, 1986; Young, 1985). The findings from this study suggest that changes in attitudes, materials, and practices are required of participants in developing achievement-level standards in programs where assessment, instruction and learning help to inform each other. Further, this implies that administrative support and staff development opportunities are critically linked to meaningful development of programs with achievement-level standards. Clearly, the process of implementing an innovation has a technical as well as a management of change dimension: there is a need to know both what needs to be done and how such change can be implemented. Although closely related, these two developmental processes are conceptually and practically distinct. Together, they provide a rationale for the merging of assessment with curriculum for the benefit of the school program, and student alike.

Teachers from the early and advanced Benchmark schools in this study all expressed a strong desire to develop their own instructional programs. Ultimately, all teachers wanted to integrate the Benchmark Program with their own patterns of instruction. The implication here is that the focus in improving classroom assessment may best be switched from implementing entire programs to providing classroom teachers with more expertise and resources in developing their classroom instructional skills. Such an approach to classroom improvement may assist teachers in integrating those assessment and instructional practices most conducive to student performance. The
approach would need to recognize that innovations may be "adapted" more than they are "adopted" for classroom use. Consequently, it may be the case that educational reformers may more often "renovate" than "innovate" programs focusing on assessment and curriculum practices.

This study has revealed the relevance of antecedents to what is changed, and how it becomes changed. Curriculum change has been viewed in terms of the interactions between its antecedents, transactions, and its outcomes. Curriculum use has been linked closely to curriculum development. Perhaps the most significant [and unstated] outcome from the variables associated with the antecedents of change concern the degree to which assessment and curriculum may be seamlessly integrated into a course of studies. The Toronto Benchmark Program is an assessment program just as much as it is a curriculum program, and vice versa. As a result, the interaction between the assessment and the curriculum itself is reciprocal, and focused on the continued development of students' classroom performances.

When we talk about developing innovative programs in our schools, we usually need to address a professional development concern. Teachers, and administrators need time and theoretical support in such things as conceptualizing the focus of the new program, its purposes and practices, its expectations, and its materials that may need to be developed. For example, ongoing teacher collaborations and inservices have helped teachers in the Montrose research site in this regard. During these collaborative sessions, the participants were able to come to understand, pilot test, discuss, and modify their classroom practices concerning the assessment innovation. These same activities were not part of the adoption process at Arcadia: teachers in Arcadia received very few professional development sessions from their Board or their school tailored to their understanding and use of the Benchmark
Program. Under normal circumstances, Black (1996) questions whether teachers can develop the skills of formative assessment, i.e., can learn to incorporate such assessment in their day-to-day classroom practice so as to improve learning without creating extra burdens for themselves. Without the accompanying professional development support, teachers' capacities to handle such roles may be very limited.

The Montrose research site had access to resources in the form of people, time and money that supported the merging of assessment and curriculum activities. Participants in Arcadia, because of the lower socio-economic status of the neighborhood did not have access to the same resources, and the benefits as did those in Montrose. It is conceivable that the participants in Arcadia needed more resources to develop and implement its assessment innovation due to the effects from its less affluent demographic environment. This factor has strong implications for the importance and logistics involved in establishing "a level playing field" for all of the schools in a school board.

In addition, the issue of autonomy in developing an innovative program is closely linked to its success in use. Within Boards of Education, program development tends to be "top-down", i.e., with development centrally controlled from the Board level with use of the program dependent on teachers accepting [at face value] the innovation into their classrooms. The Montrose research site was very fortunate in that its school administration was involved in the conceptualization, design, and development of the Toronto Benchmark Program. Teachers at Montrose had immediate contact with the program's founders, and also the benefit of their expertise in the two years they spent preparing to use the Benchmark Program in their classrooms. In effect, the Toronto Board of Education took its lead from the efforts of the administrators, and teachers at Montrose with the Program's objectives,
activities, materials and assessment instruments. The administrators and teachers at Arcadia did not have much opportunity to develop the Benchmark Program at their local school level.

Also worthy of mention is the decision-making and leadership support each research site received from their respective school administrators. A very critical incident occurred at Montrose when the administrator introducing the Benchmark Program to his staff decided *for his staff* that they would be participating in the development and use of the Program. That is, Montrose teachers were asked to attend school noon-hour workshops on the Benchmark Program every two weeks, and were encouraged to use more than the Board-mandated two Benchmarks in their classrooms during the school year. In comparison, the Arcadia administrator communicated to her teachers the Board mandate that each teacher use two Benchmark lessons during the school year, but no regular school workshops on the Benchmarks were scheduled for teachers at Arcadia. At Montrose, the development and use of the Toronto Benchmark Program was an administrative as well as a classroom priority; at Arcadia it was not.

Missing from both research sites was data on the critical issue of the school-level context for curriculum decision-making. Inherent in developing and using any innovation is the capacity to make decisions that will promote the successful implementations of innovations into school classrooms. Although both Montrose and Arcadia schools had mission statements displayed on their walls, these statements were not used as contextual screens for their curriculum decision-making. Further, no other statements or policies were available for use in curriculum decision-making. The infrastructures of each were not equal in terms of access to development time, educational expertise, and finances. Montrose teachers benefitted from the extraordinary resources
available to them. Teachers at Arcadia, for the most part, were left unaware of the meanings, purposes, and potential benefits of merging curriculum with assessment.

The issue of curricular orientation needs to be considered in this study. The Benchmark innovation was structured in terms of its objectives, standards, and outcomes. The instructional process was prescribed for teachers in somewhat of a mechanistic linearity, i.e., it was consistent with the technological orientation to curriculum (Eisner & Vallance, 1974) in which classroom activities were designed to achieve pre-specified ends. Consistent with this orientation, the Language Arts curricula were analyzed into its component parts, i.e., atomistically. It is possible that such a program runs the risk of 'boxing in' teachers, i.e., forcing teachers to become instrumental data gatherers more than participants in the educative process. For example, Torrance (1995a) explains how researchers McCallum, Gipps, McAlister & Brown documented variations amongst teachers in 32 schools (with case studies in six of the schools) in their perspectives on and approaches to assessment. They summarized the variations in terms of three idealized types which they called "intuitives", "evidence gatherers", and "systematic planners". Only the 'systematic planners' made a close link between assessment and instruction. They not only planned assessment activities on a weekly basis, but also used the results in planning instruction, often using the results for diagnostic purposes. Both the 'intuitives' and 'evidence gatherers' did not integrate their assessment with their teaching activities: both were kept separate from each other. This study by McCallum et al. (1995a) offers the potential to proceed with systematic inquiry designed to help clarify how the 'systematic planners', as individuals and as a group, were able to adopt, and
manage their assessment innovation so well. *How did their development and implementation practices compare to the intuitives, and the evidence gatherers? How are the systematic planners able to continue to practice the innovation in their classrooms? How do the instruction, assessment and learning experiences in the classrooms of the systematic planners compare to those of the others?* Data related to these questions might be helpful in clarifying the supports and practices conducive to developing and managing assessment programs.

The recent introductions of the *Common Curriculum* (1993), and province-wide testing (EQAO, 1997) in Ontario not only represent a shift to an assessment and standards-based emphasis on curriculum, but also points to the fragile balance between assessment and curriculum. For example, in developing and providing curriculum and testing for the teachers and students in Ontario's schools, the Province of Ontario (not the teachers and students at the local school level) is prescribing for teachers and students what knowledge they must adopt. What initially may set out to be well-intentioned curriculum-driven assessment could easily evolve into an unofficial, but nevertheless, an operational assessment-driven curriculum. Because the imbalance between CDA and ADC can have consequences deleterious to student performance, teachers, students, administrators and parents need to monitor the process of curriculum development and implementation, and also identify how the curriculum is being driven. The essential questions to be concerned about are *Who is doing the driving?*, *Why it is being driven?*, and *What are the consequences of the driving?*.

Studies of classroom assessment need to move beyond the evaluation process, e.g., frequency of testing, towards more comprehensive inquiries
involving tasks, criteria, standards, samples of information on performance and outcomes, appraisals and feedback. In turn, teachers confronted with the demands of the classroom will need to consider the full range of issues in developing an approach to student evaluation. In this case, both research and practice will be improved by adopting a more comprehensive framework for studies of classroom assessment.

More than a few studies seek to develop alternatives to norm-referenced standards, but descriptive accounts suggest that such standards may not be used extensively by teachers at the present time. Additional research is needed to provide a fuller account of how students are currently evaluated in classrooms with respect to criterion-referenced standards. This research will need to consider explicitly which of the multiple purposes of assessment can be served by which combinations of practices. As a result, many studies compare different assessments methods in terms of some outcome that may have nothing to do with the purpose for which the methods were developed. Studies will also need to link the characteristics of the context of each classroom in terms of which assessment demands are instrumental in practice, and how they influence curriculum and assessment.

A common theme in this study has been that teachers develop a style of instruction that is in line with their own philosophies of education; styles that are independent of any innovation that they may be expected to develop and use. In all eight cases of teachers in this study, teachers were not willing to compromise their teaching style for one that was foreign to them. This situation is ripe for research that seeks to clarify whether instructional development needs to precede program development, and how teachers' unique instructional styles can be merged with implementing an innovation into their classrooms.
This study has also revealed that in the content-based classrooms, the mainstay of instruction was materials from which knowledge could be imparted to students. Research might be developed that would clarify the relationship of teachers' materials with lesson objectives, and styles of instruction. For example, *To what extent do the materials adopted become the actual curriculum?*, *What is the nature of assessment of the knowledge in these materials?*, and *How do the materials integrate with the assessment practices?*.

Since instruction and assessment practices help to compose a major part of the teaching function, information and training for teachers and students on methods of integrating assessment with instruction might be well received, particularly at the preservice level. For example, such training sessions could focus on 1) developing a table of final test specifications for a final unit before it is taught; 2) involving students in writing practice test items so that students will have to evaluate the importance of the various elements of content; 3) and using lists of unit objectives and tables of test specifications to communicate with other teachers about instructional priorities, and to arrive at a clearer understanding of how these priorities fit together across instructional levels, schools, and an entire Board. Another component of such workshops might involve the steps involved in developing performance criteria for one's own classroom. For example, the sessions could focus on the outcomes expected from a particular curriculum, and teachers' [tacit knowledge] could be consulted for the standards of that particular curriculum. These standards could then be articulated and written down with indicators added to make them more explicit. To determine the indicators or criteria, teachers could refer to samples of student work or to norm-referenced test results as a basis for measuring performance over time.
Further research is needed to develop an understanding of how systems of meaning that support current assessment practices within school cultures developed, and what factors may influence changes in them. For the participants at Montrose, the Benchmark Program was an innovation that provided a necessary link in the "chain of development" at that school. Montrose teachers began with an interest in developmental learning, moved to whole language instruction (which emphasizes teacher-student interaction, and the integration of course content), to the Benchmark innovation which merged formative with summative assessment, and to a system of portfolios designed to help record, and report on their students' classroom performances. The Benchmark innovation contributed to the Montrose teachers' ongoing professional growth. The teachers from Arcadia were developing their classroom programs on an individual basis with the goal of establishing an instructional protocol with which each felt comfortable. However, the Montrose experience has reinforced that teachers, administrators and researchers, when working together to analyze relationships that exist with individually constructed beliefs in assessment and instruction, may come to question the rationale for their assessment practices, and then may be more prepared to modify them in order to enhance the classroom performances of their students.
Bibliography


Toronto Board of Education (1979). *Evaluation, measurement and reporting to parents: The report of the staff committee on evaluation and testing.* Toronto: Curriculum and Program Division.


Toronto: Toronto Board of Education.


Appendix A

Teacher Interview Schedule

1. What do the terms i) 'assessment', and ii) the 'Toronto Benchmark Program', mean to you?

2. What would you see to be the optimal relationship between assessment and curriculum in your course? (Please explain.)

3. i) For how long have you been developing the Benchmark Program in your school? 
   ii) What is the basis for your involvement in the Program? (e.g., volunteer, assignment)

4. Could you describe what you think constitutes your assessment and curriculum development activities within the Benchmark Program?

5. How would you compare your experience with assessment and curriculum development in the Benchmark Program to any non-Benchmark Program development experience?

6. Can you articulate the assessment policy for your course? (How was it developed? How does it work? How do you and your students fit into it?)

7. What types of assessments do you use with your students?
   i) How frequently do you use them?
   ii) What does the data from the assessments tell you regarding your students and your course?
   iii) In what ways do you use this data?
   iv) Are you satisfied with the utility of this data? (Or would you prefer to use data in different ways, and from different kinds of assessments? Please explain.)

8. How would you see the relationship between assessment and your curriculum now as compared to that in former (non-Benchmark Program) curricula? (e.g., assessment as driving the curriculum or curriculum as driving your assessment of students)

9. i) Based on your experience and opinion, what are the benefits and weaknesses of each of these assessment-curriculum relationships on i) your curriculum, i.e., educational activities in your classes, e.g., instruction, and ii) school practises, e.g., parent-teacher meetings?
   ii) Which relationship do you prefer? (Or would like to see changed - how and why?)
Appendix B

Student Interview Schedule

1. i) What do you think testing is? Could you explain it to me?
   ii) Do you know what the Toronto Benchmark Program is? Could you explain it to me?

2. Would you like to have more tests (so you would know how you are doing in each subject) or would you rather have more activities in class? Why?

3. i) Is this your first year in this grade?
   ii) What do you like about the way you are i) tested and ii) taught this year?
   iii) Is there anything you don't like about the way you are i) tested and ii) taught this year?
   iv) Were i) your class activities, and ii) your testing different for you last year? How?

4. i) Does your teacher ask you which activities you like best, and what questions you like to do?
   ii) Does your teacher use your suggestions for activities?

5. Has your teacher told you how and why you are being tested? What did she/he tell you?

6. What kinds of tests (pencil-and-paper, oral) does your teacher give you?
   i) How often do you have to do them?
   ii) What do these tests tell you about yourself and your course?
   iii) Does that help you in any way? How?
   iv) Do you think your tests are fair? Could you tell me how or why?

7. i) Do you think that what you did in class helped you with the tests?
   ii) Which do you think was more important - classroom activities or tests? Why?

8. i) Do your parents/guardians know what you are doing in class? How well you are doing?
   ii) How do they know? (Report cards, talking to the teacher, newsletters, seeing you in class?)
Appendix C

Administrator Interview Schedule

1. What do the terms i) 'assessment', and ii) the 'Toronto Benchmark Program', mean to you?

2. What would you see to be the optimal relationship between assessment and curriculum in your teachers' classrooms, i.e., where might the emphasis be placed for best educational results? Please explain.

3. i) For how long have you been developing the Benchmark Program in your school?
   ii) What is the basis for your involvement in the Program? (e.g., volunteer, assignment)

4. Could you describe what you think constitutes your assessment and curriculum development activities within the Benchmark Program?

5. How would you compare your experience with assessment and curriculum development in the Benchmark Program with any non-Benchmark Program development experience?

6. Can you identify the assessment policy for the courses delivered in your school? (How was it developed? How does it work? How do you, your teachers, and their students fit into it?)

7. i) What data do you see (expect) generated by the assessments in your teachers' classes?
   ii) How do you see (expect) the data from these assessments to be used in teachers' classes?
   iii) Are you satisfied with the utility of this information? (Or would you prefer to use data in different ways, and from different kinds of assessments? Please explain.)

8. Generally, how do you generally see the relationship between assessment and curriculum currently in your school as compared to that in former non-Benchmark Program curricula? (e.g., assessment as driving the curriculum or curriculum as driving assessment of students)

9. i) Based on your experience and opinion, what are the benefits and weaknesses of each of these assessment-curriculum relationships on i) the curriculum, i.e., educational activities, in your teachers' classes, e.g., instruction, and ii) school practices, e.g., reporting student progress to parents?
   ii) Which relationship do you prefer? (Or would like to see changed - how and why?)
Appendix D

Parent Interview Schedule

1. (a) What is your understanding of
   i. Benchmark Program, and ii. curriculum, e.g., as classroom activities?
   (b) Whose interests do you think are being served through Benchmark portfolio assessment?
      (e.g., students, teachers, administrators, parents) How?
   (c) To your knowledge, what is the current purpose of Benchmark portfolio assessment in your
      child's class?
   (d) What do you think should you be the main purpose(s) of Benchmark portfolio assessment?

2. (a) Have you had any input or involvement in the development of the Benchmark portfolio
      assessment your child receives in his school?
   (b) Are you currently involved in the development of the Benchmark portfolio assessments in
      any way? (e.g., community and school conferences)
   (c) What do you think have been (or are) the major issues involved in developing the
      Benchmark portfolio assessments to be used in your child's classes?

3. (a) How would you describe the way in which Benchmark portfolio assessments are used with
      your child in his/her class?
   (b) How well do you think these assessments show you how your child is doing in class?
   (c) Do you use this information that is reported to you in any way? Could you explain?
   (d) As a parent are you satisfied with the usefulness of this information?

4. (a) When looking at your child's subject material, and his/her assessment, do you think:
      i. one should be more important than the other? Which one, and why?
      ii. both are just as important, and should work together. Why, and how?

5. (a) In terms of your child, do you think there are any benefits and/or weaknesses of Benchmark
      portfolio assessment and/or how your child is being taught? (e.g., reporting your child's
      achievement to you).
   (b) How does this compare to that when your child was assessed with pre-Benchmark
      assessment and instruction?
Appendix E

Classroom Observation Schedule

1. i) What characteristics of the particular classroom setting relate to the curriculum and its assessment of students?
   ii) What appear to be the dynamics in the assessment-curriculum relationship for the classroom observed?

2. a) How is the particular course material taught and assessed?
   b) What do the students seem to comprehend by the instruction and assessment?
   c) Do the instruction and assessment appear to be meaningful for the teacher and students?

3. What are the effects of this relationship on the educational activities of the classroom observed and/or practises of the school?
Appendix F

Toronto Board of Education Informed Consent

Dear Research Manager,
Toronto Board of Education

I am writing to request approval from the Review Research Committee to conduct my thesis study in your school system. This research will help me complete requirements for my Ph.D. in Curriculum Studies at the Ontario Institute for Studies in Education.

The purpose of my study is to explore the nature of the assessment-curriculum relationship and its consequences on school curricula and practice by examining the perspectives of the teachers, students and administrators associated with the Toronto Board Benchmark Program. Data collected will help to provide educators with knowledge of a curriculum's actual potential for delivering on its learning, instructional and accountability goals within the contexts of its local site.

I am requesting access to two schools - one in which the Toronto Benchmark Program has been developed, and the other, in the early stages of implementation of its Grades 3, 6 and 8 Benchmark language and math classes. In each school, I will need to conduct 45-minute semi-structured interviews and participant observations with six teachers, six students, and one school administrator directly involved with the Benchmark Program. I anticipate the data collection process to take place sometime between October of 1993 and January of 1994.

This study employs no deceptions or perceived negative effects for any of its participants. Their identities will be kept anonymous, and their responses confidential through the coding, non-discussion, and secure storage of data. Thus, the data will not be used to evaluate teacher, student or administrator performance. All participants will be free to make inquiries about this study by contacting me at 949-0969, or my thesis advisor - Dr. Joel Weiss at 923-6641. Should they choose, they would be free to withdraw from the study at any time, without penalty.

Enclosed are two copies of this letter. If the Research Review Committee decides to approve my application, please sign your name under "Signature of Consent", and mail that copy to me in the self-addressed envelop provided. The remaining copy is for your records.

Thank you for considering my research request.

Signature of Consent, Yours truly,
Research Manager Ron Saranchuk
Appendix G

Administrator Informed Consent

Dear Principal,

___________ School,

Toronto Board of Education

The Review Research Committee has approved my request for conducting my doctoral thesis research within the Toronto Board. However, I am also requesting your approval to conduct research in your school. I will need to interview you, and interview and observe (as a participant) six teachers and six of their students in your school who are using the Toronto Benchmark Program in their Grades 3, 6 and 8 language and math classes. I anticipate this data collection to take place sometime between October, 1993 and January, 1994.

The purpose of my study is to explore the nature of the assessment-curriculum relationship and its consequences on school curricula and practice by examining the perspectives of the teachers, students and administrators associated with the Toronto Board Benchmark Program. Data collected will help to provide educators with knowledge of a curriculum’s actual potential for delivering on its learning, instructional and accountability goals within the contexts of its local site.

This study employs no deceptions or perceived negative effects for any of its participants. Their identities will be kept anonymous, and their responses confidential through the coding, non-discussion, and secure storage of data. Thus, the data will not be used to evaluate teacher, student or administrator performance. All participants will be free to make inquiries about this study by contacting me at 949-0969, or my thesis advisor - Dr. Joel Weiss at 923-6641. You, your teachers, and their students would be free to withdraw from the study at any time, without penalty.

Enclosed are two copies of this letter. If you approve my research request, please sign your name under "Signature of Consent", and mail that copy to me in the self-addressed envelop provided. The remaining copy is for your records.

Thank you for considering my research request.

Signature of Consent, ____________________________

Yours truly,

School Principal ____________________________

Ron Saranchuk
Appendix H

Teacher Informed Consent

Dear Classroom Teacher,

__________________________School,
Toronto Board of Education

The Review Research Committee has approved my request for conducting my doctoral thesis research within the Toronto Board. As well, your principal has given me permission to conduct my research in your school. However, I am also requesting your approval to interview and observe you (as a participant), and two of your students in your Grade ______ Toronto Benchmark language/maths class. I anticipate this data collection to take place sometime between October, 1993 and January, 1994.

The purpose of my study is to explore the nature of the assessment-curriculum relationship and its consequences on school curricula and practice by examining the perspectives of the teachers, students and administrators associated with the Toronto Board Benchmark Program. Data collected will help to provide educators with knowledge of a curriculum's actual potential for delivering on its learning, instructional and accountability goals within the contexts of its local site.

This study employs no deceptions or perceived negative effects for any of its participants. Their identities will be kept anonymous, and their responses confidential through the coding, non-discussion, and secure storage of data. Thus, the data will not be used to evaluate teacher, student or administrator performance. All participants will be free to make inquiries about this study by contacting me at 949-0969, or my thesis advisor - Dr. Joel Weiss at 923-6641. You and/or your students would be free to withdraw from the study at any time, without penalty.

Enclosed are two copies of this letter. If you agree to participate in my research study, please sign your name under "Signature of Consent", and mail that copy to me in the self-addressed envelop provided. The remaining copy is for your records.

Thank you for considering my research request.

Signature of Consent, \hspace{1cm} Yours truly,

Classroom Teacher \hspace{1cm} Ron Saranchuk
Appendix I

Parental Informed Consent

Dear Parent,

I am conducting a study to find out how young students are affected by the relationship between classroom subjects and how they are tested. I would like to include your child in my study.

I will arrange with the classroom teacher so that I can talk to your child at school during a time when he/she will not be missing any school work. I will ask them questions about their subject and how they are tested, e.g., "What kinds of questions do you do in class?" I would like to speak to your child once, for about 30 minutes, between October of 1993 and January of 1994. I will keep confidential what your child tells me. As well, I will neither ask for nor use any information about your child from the school's records.

This study has been approved by the Research Review Committee of the Toronto Board of Education, your child's school principal, and your child's classroom teacher. This research is important because it help to show how classroom activities and testing might be changed to help your child learn in school.

Please complete the form at the bottom of this letter and return it to your child's teacher by ___________. Your child will be free to withdraw from this study at any time, without being penalized. If you would like to receive more information about the study, please phone me at 949-0969, or my research supervisor (Dr. Joel Weiss) at 923-6641.

Thank you,

Ron Saranchuk

---------------------------------------------------------------

Child's name____________________________________________________
Birthdate____________________ School___________________________

CHECK HERE
______ I give permission for my child to participate.
______ I do NOT give permission for my child to participate.

Signature of parent/guardian____________________________________

---------------------------------------------------------------
Appendix J

Toronto Benchmark Performance Criteria

Performance Criteria (L3). Upon observation of student performances with the sixteen grade three Language Arts Benchmarks, the following list of qualities were repeatedly displayed amongst students who achieved well:

- Interpreting
- Expressing opinions
- Drawing inferences
- Elaborating and extending ideas
- Self-monitoring
- Checking thoughtfully
- Questioning and seeking clarification
- Taking initiative
- Comparing and evaluating
- Predicting
- Speculating
- Synthesizing
- Revealing humour and imagination
- Showing appreciation
- Demonstrating interest and commitment
- Taking risks
- Summarizing
- Paraphrasing
IMAGE EVALUATION
TEST TARGET (QA-3)

APPLIED IMAGE, Inc.
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved