

## **INFORMATION TO USERS**

**This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.**

**The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.**

**In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.**

**Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.**

**Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.**

**Bell & Howell Information and Learning  
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA  
800-521-0600**

**UMI<sup>®</sup>**



## **NOTE TO USERS**

**This reproduction is the best copy available**

**UMI**





**Université d'Ottawa • University of Ottawa**



**Exploring the Technical Qualities of the Self-Assessment Process  
in the Intermediate Grades**

© Leslie McLeod

Faculty of Education

University of Ottawa

Thesis Supervisor: Professor Marielle Simon

Thesis Committee: Professor Hanne McWhinney

Professor Renée Forgette-Giroux

March 1999



National Library  
of Canada

Acquisitions and  
Bibliographic Services

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque nationale  
du Canada

Acquisitions et  
services bibliographiques

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*

*Our file* *Notre référence*

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.

0-612-45240-9

Canada

## ABSTRACT

Today's rapidly changing, technological society demands that educational bodies produce individuals who are lifelong learners. However, beyond the traditional schooling years, individuals seldom have the opportunity to learn within a formal setting. Teaching individuals to self-assess is a logical adjunct to the goal of producing lifelong learners. This focus of this study is to explore the implementation of a proposed self-assessment framework in grades 6, 7 and 8. The described process is closely examined to determine whether the results produced are both valid and reliable.

Educational assessment literature proposes a number of advantages regarding the use of self-assessment. Assessment experts recommend that it provides an excellent way of familiarizing the students with the proposed learning standards as both students and teacher must have similar understandings of the assessment criteria. Others suggest that students are also able to participate more responsibly and enthusiastically in the learning process if they have a deeper understanding of the learning goals and related standards for performance.

Although self-assessment is recommended with enthusiasm in the professional teaching literature, there is little scientific research that supports its use as an assessment process in the elementary grades. The present study is exploratory in nature and examines the self-assessment process by investigating the implementation of a researcher-proposed framework for student self-assessment.

Six teachers participated in this study and worked with their students to develop rubrics that were used for student self-assessment in paragraph writing. The framework included implementational steps as well as checks to ensure the validity and reliability of the process. The self-assessment process undergone by these six classes is documented and the assessment results are analyzed to determine levels of validity and reliability. In addition, the researcher has identified important factors which can impact the validity and reliability of the results.

The findings of the study are presented in this thesis and possible directions for further research have been identified.

## **Acknowledgment**

I would especially like to thank my thesis supervisor Professor Marielle Simon for her guidance and encouragement throughout this project. She has helped to make the journey an interesting and enjoyable one. I also want to thank Professors Renee Forgette-Giroux and Hanne Mawhinney for their insightful suggestions and interest in my work. Lastly, but most importantly, I thank my husband Hans for his support; both emotional and financial and my cat Barney for keeping me company during all the lonely hours at my computer.

## **Table of contents**

1. Abstract

2. CHAPTER 1: Introduction/8

3. CHAPTER 2: Literature review/12

    Merits of Self-Assessment/13

    Implementation Issues in Self-Assessment/15

    Technical Issues in Self-Assessment/21

    Concepts and Issues Relating to Present Study/28

        A) Research Questions/28

        B) Definition of Terms/28

        C) Self-Assessment Conceptual Framework/31

CHAPTER 3: Method/38

    Participants/38

        Teachers and Students/38

        Consultants/41

        Researcher Background/41

    Instrumentation/42

    Procedure/44

Step One: Intention/46

Step Two: Measurement/47

Step Three: Measurement and Judgment/48

Step Four: Debriefing/49

Analysis/51

## CHAPTER 4: Results/53

The Self-Assessment Process/53

External Process and technical checks/54

Analysis of Self-Assessment Data/60

Internal Process/61

Similarities and Differences Between Class Groups/68

Factors Affecting Technical Quality /69

Achieving a Common Understanding Across Ability Levels/71

Rubric Development and Practice/72

Students' Attitudes Towards the Self-assessment Process and Student Ownership/74

## CHAPTER 5: Discussion/77

Limitations/85

Implications for Future Research/86

Conclusion/88

REFERENCES/90

## **List of Figures and Tables**

Figure 1- Student self-assessment conceptual framework/37

Figure 2- Overview of study timeline/46

Figure 3- Implementation framework outlining external and internal aspects of the process/50

Figure 4-Modified student self-assessment conceptual framework/84

Table 1-Teacher participants/40

Table 2- Overall inter-rater agreement/61

Table 3- Inter-rater agreement within classes/61

## **APPENDICES**

**Appendix A: Data Collection Instruments**

**Appendix B: Final Version of Self-Assessment Rubrics**

**Appendix C: Synopses of Internal Process Within Each Classroom**

**Appendix D: Exemplars**



## Chapter 1

### Introduction

Recent calls for educational reform have urged schools to respond to the needs of our rapidly changing society and to produce graduates who not only have the necessary basic skills for the workplace but are critical thinkers, creative problem solvers and self-directed learners. As early as 1980, the Ontario Ministry of Education described the educated person as a self-directed problem solver, a methodical thinker who is capable of inquiry, analysis, synthesis and evaluation (Ministry of Education, Ontario, 1980).

As a result of these demands, a number of alternative methods of teaching and assessment have gained widespread acceptance during the latter decades. Traditionally, instruction endeavoured to impart specific knowledge units to learners and to assess student mastery regarding these units. Spady and Marshall (1991), have argued that conventional learning goals are no longer the only answer and that educators must broaden their targets to include goals that relate to life-long learning. An important aspect of many of these described goals, is the students' ability to assess themselves. Beyond the traditional schooling years, individuals seldom have the opportunity to learn within a formal setting. Self-assessment's main attraction is its potential to

help learners set individual goals and measure their progress in accordance with these goals. In essence, if educators are to reach their goal of producing lifelong learners, self-assessment is a vital skill.

One of the main emphases of this educational perspective is that students and teachers act as collaborators in the learning process (Spady & Marshall, 1991). Research has demonstrated that student achievement is enhanced when the learner is involved in the assessment process and understands learning goals (Glaser, 1990, Sadler, 1989). Wiggins (1993) suggests that students cannot achieve a goal effectively, if they are unaware of achievement standards or the specific achievement task is not communicated in advance. Self-assessment is well suited to this perspective on learning as an individual must be aware of the learning objectives as well as the standards for performance if self-assessment is to be accurate. The essential notion supporting self-assessment is that it provides both the teacher and the learner with a better understanding of how the learner learns, as well as an awareness of areas of learning yet to be mastered (Black & Broadfoot, 1982; Wiggins, 1993.) Common sense suggests that individuals are best acquainted with certain aspects of their learning such as effort or personal attainment goals. From this standpoint, self-assessment can only enlarge the assessment picture for both teacher and learner. Lastly,

self-assessment is respectful of learners because it allows the learner to be a stakeholder in the assessment process, within a variety of learning opportunities.

Self-assessment is currently receiving increasing emphasis in school systems, with school boards asking teachers to incorporate self-assessment into their existing assessment schemes. However, teachers are being asked to implement it with little knowledge as to appropriate self-assessment methods or as to the technical qualities of self-assessment as an assessment process (Satterly, 1989; Gipps, 1994). At the present time, the benefits of self-assessment remain largely theoretical and little research has been done which supports the use of self-assessment in the school classroom. Instead, research regarding self assessment has been focused for the most part on adult learners (Boud and Falchikov, 1989). Recent studies indicate a growing interest in self-assessment, but a common theoretical focus or practical framework is lacking.

The purpose of this study is to explore many of the issues highlighted here by investigating the use of a self-assessment process in classes ranging from grades six to eight. In particular, it examines to what extent teachers and students can work together to implement a valid and reliable process for self-assessment within the classroom for formative ends. A proposed framework to

guide the implementation was developed by the researcher as a means of empowering teachers to use the self-assessment process with their students.

The subsequent review of the literature presents the current body of research focusing on self-assessment; its merits, as well as research regarding implementational and technical issues. The third chapter describes the both the external (events occurring outside the classroom) and internal (events occurring within each classroom) processes in which teachers and students participated to implement the proposed framework for self-assessment as well as the researcher's role in the intervention. Chapter four presents the implementation process as it occurred and describes the external and internal steps taken to ensure the validity and reliability of that process. Chapter five discusses the relevance and limitations of these findings as well as further implications for future research in self-assessment.

## Chapter 2

### Literature Review

The following literature review is divided into four sections. The first section discusses the merits of self-assessment as they have been described by educational field practitioners and researchers. The second and third sections explore the implementation and technical issues that have been identified by a number of studies focusing on self-assessment. Each of these three sections highlights different perspectives of the self-assessment process. The reader will note that the findings of the presented studies may be discussed in more than one section. This is due to the relevance of studies' findings to the three sections presented here as well as the scarcity of research relating to the technical qualities of self assessment. The final section focuses on concepts and issues relating to the present study and presents the specific research questions which arose from the literature review to guide this study, a description of terms as they are used in this study, and lastly the researcher's proposed conceptual self-assessment framework.

### Merits of Self-Assessment

In self-assessment, the learner occupies a central position and is thus responsible for reflecting on and evaluating his or her achievement level with respect to set standards (Satterly, 1989). In order to do this effectively, performance expectations must be clearly stated and standards or criteria for performance must be developed and understood by all users. Once this has been done, then performance can be judged in accordance with these standards (Boud & Falchikov, 1989). According to Boud and Falchikov, (1989), who undertook what appears to be the most comprehensive meta-analysis to date of the literature on student self-assessment in higher education, this type of assessment is adaptable to both formative and summative assessment. Its flexibility also allows the learner to choose and clearly state what aspects of achievement are to be assessed. In turn, these may be global, complex skill acquisition or one small component of a larger learning goal.

Both scientific and professional educational literature propose a number of advantages regarding the use of self-assessment. According to Black and Broadfoot, (1982), it provides an excellent way of familiarizing the students with the proposed learning standards as both students and teacher must have similar understandings of the assessment criteria. Gipps (1994) states that the

students' understanding of the learning criteria is a key aspect of formative self-assessment. Gillen (1991,) suggests that students are also able to participate more responsibly and enthusiastically in the learning experience if they have a deeper understanding of the learning goals and related standards for performance. When learning is self-regulated, the student can have a greater impact on the assessment and is therefore, more able to take greater ownership for learning as it occurs (Mckinney, Perkins & Jones, 1995.) In other words, as students take over ownership for learning, they are able to set their own goals and choose the means for documenting their work towards these goals (Mckinney et al, 1995.) As a result of this, the individual learns to develop criteria for assessment which can then be applied to diagnose strengths and weaknesses in order to refine and adjust learning tasks along the way. This more active and autonomous involvement on the part of the student also helps students to understand the reasons for their own educational performance. Given that, once the learner exits school there is little opportunity to be assessed by an outside body, learning how to self-assess allows one the opportunity to provide formative feedback for one's self, regardless of the learning environment.

Despite these merits, questions regarding the feasibility and technical soundness of self-assessment remain largely unanswered. Boud and Falchikov (1989), highlight the need for future study including: exploring ways to maximize student and expert agreement in order to strengthen evaluation reliability, searching out areas in which student assessment can be used effectively, as well as discovering the key variables that influence self-assessment outcomes. Without answering some of these important research questions and establishing its soundness as a means of student evaluation, self-assessment is likely to be a short lived trend.

The following section discusses a number of applications of different self-assessment processes that have been developed. For the purpose of this review, the studies being discussed have been divided into two categories: Implementation issues and Technical issues.

### Implementation Issues in Self-Assessment

According to Satterly (1989), psychological theory suggests that children develop a greater understanding of self with increasing age, experience and cognitive development. Perhaps it is this reason that accounts for a greater

attention to self-assessment research in post secondary education as is evident by the studies reported here.

A doctoral study by Anderson (1993) looked at self assessment in the university undergraduate English classroom. The research question focused on the conditions under which freshman English students develop and accurately use self-assessment for improving their work (Anderson, 1993). Although a general rubric identifying and describing assessment criteria was distributed at the beginning of the course, no specific process for self-assessment was taught to the students involved in the study. Students participated in one on one tutoring sessions with a course teaching assistant to work at assessing and revising their writing. One of Anderson's major findings emphasized the students' need to understand the instructor's criteria for grading in order to apply similar criteria in their self-assessment. Providing students with a rubric outlining the assessment criteria was not enough to ensure a common interpretation between students and instructor. Exemplars must be shared with the students to illustrate a range of achievement and discussed with the whole class in relation to the rubrics in order to increase the students' understandings of the evaluation criteria (Anderson, 1993). A second finding of this study was that students of different capability levels adopted the self-assessment process

at different rates and at different levels of complexity (Anderson, 1993). For example, students who were weaker writers were less adept at self-assessment than stronger ones. No data suggesting a basis for this finding was provided.

Long and Fransen (1986) investigated what they termed "the trustworthiness" of self-assessment with professionals participating in a learning conference. A Likert scale type survey was given to the participants before the conference in order to identify their learning objectives for the conference and once again when it was over to determine whether these objectives were met. The survey data was then triangulated with senior managers who were considered knowledgeable experts about their employees performance. A number of conditions were put in place which were believed would increase the trustworthiness of the self-assessments including: criterion achievement scales based on behaviourally oriented learning objectives and described in language that promotes common understanding, attribution data were gathered about learning activities that helped explain the changes in achievement ratings, suggestions for future learning stemmed from the self-assessments just completed, and self-assessment perceptions are triangulated with perceptions of other knowledgeable experts (Long & Fransen, 1986). Two interesting findings that arose from the study are the need for explicit criteria

in order to increase the accuracy of the results as well as triangulation in order to assess the reliability of the process.

A third qualitative study looked at the use of self-assessment portfolios in an education literacy methods course and discussed its value as an assessment tool (McKinney, Perkins & Jones, 1995). During the first two weeks of the course, student teachers were required to submit a list of goals that they wanted to achieve and ideas about how they would provide evidence to indicate that they were meeting these goals. During the term they were asked to generate characteristics of "A work" and "not A" work (McKinney et al., 1995, p.23). These characteristics were then shaped to correspond with the goals stated in the university course syllabus. Students were required to collect samples of their work throughout the term as well as reflections on this work and their practice. When the project was complete, both students and two external reviewers of the portfolios (college professors teaching other sections of the course) were asked to analyze the value of the self-assessment portfolio and determine whether it could stand on its own as a sufficient indicator of learning progress. Both groups felt that there were a number of positive impacts of the self-assessment portfolios but also suggested there may be a need for multiple indicators of achievement and suggested that test scores

should be included in the assessment (1995). The authors suggest that these responses might relate to the feeling on the part of the participants of a need for greater clarity as to what exactly is to be assessed and the criteria for that assessment, underlining once again the need for a common understanding of explicit criteria (1995).

Interestingly, in a study focusing on elementary school children, Higgins, Harris and Kuehn investigated first and second grade students' generation of assessment criteria and their self-assessment of projects using these criteria (1994). Rather than presenting the students with a pre-determined rubric to use as a judgment tool, the students and teachers participated together to develop criteria which the students then applied to their work using a five point rating scale for each criteria (1 being low quality, 5-high.). The study found that contrary to popular belief, young children are capable of helping to develop assessment criteria and then choosing and applying various criteria to assess their work. Another interesting issue relating to the use of self-assessment arose from this study. Although the class developed the criteria together, when assessing their work, students chose the criteria that were most relevant to themselves for self-assessment, thus producing a tailor made interpretation framework within the assessment process. Higgins et al. suggest

that if children are to find value in the self assessment process they need to be challenged to assess themselves on criteria that they value. The results indicated that when students were given the opportunity to develop criteria for self-assessment that these criteria often reflected the teachers' learning goals for the students.

One aspect of self-assessment that is repeatedly underlined by these studies is the importance of developing criteria for performance assessment and ensuring a common understanding of these criteria. These studies also indicate that there is positive evidence supporting self-assessment as a means of improving practice or learning achievements. However, each of the studies discussed is an example of a one time isolated project and little replication has been done to build upon or connect with previous studies. Perhaps, as Boud and Falchikov (1989), suggested, it may be due to the lack of consistency in the locus of published self-assessment research which may account for this weakness. It appears that a focused conceptual framework still needs to be developed for implementing various aspects of the self-assessment process; in particular, the development of valid, commonly understood criteria.

### Technical Issues in Self-Assessment

One of the problems that becomes apparent when reading the scattered literature on self-assessment is that it has not been clearly defined as a concept or process. The various literature often focuses on the data interpretation step of the assessment process and includes a variety of forms such as portfolio self-reflections, reflective journals, and self reports under the umbrella term of self-assessment. When discussing the technical aspects of self-assessment, studies using these different interpretation methods are presented in order to highlight the evidence that supports self-assessment as a process while at the same time underlining the numerous questions which must be answered if it is to be accepted as a valued one.

For this paper, Linn and Gronlund's (1995) definition of validity will be used. In their terms, validity refers to the adequacy and level of relevance of assessment interpretations in relation to a specific use. Two validity issues should be considered important regarding assessment within the educational context. Firstly, the level of adequacy that the assessment process meets in representing the content and competencies of a learning domain about which judgments are to be made must be determined. The second important issue refers to whether the assessment yields pertinent information to consumers,

whoever they may be. These important issues can be answered in part by careful consideration of the stated achievement goal, criteria for achievement and the selection of learning tasks for data gathering (Herman & Winters, 1994).

A 1980 study by Dumont and Troelstrup using self-report data looked at the relationships between objective and subjective measures of instructional outcomes within the self assessment process. Self-report is a broad term meaning self-estimates of performance and includes methods such as survey data interpreting performance or qualitative self-reflections of learner performance (Boud & Falchikov, 1989). Specifically, the researchers' concern was the issue of using student self-report data as a form of self assessment and the validity of these assessments as performance indicators of selected educational outcomes. The student self-reports were answers to survey questions (developed by the course instructors), regarding work completed and the individual's perceived level of achievement on the assigned tasks. The self-report data demonstrated higher validity for non-academic indicators of student background and experience and low correlation with performance on standardized tests. The authors concluded by asserting that more research as to the aspects of learning in which self-reporting can be done and as to whether it

can be considered as a valid measure of student achievement must be conducted.

A review of the research done by Pike (1995), regarding the use of self-reports again as a form of data gathering for the self-assessment process raised two other interesting issues. Firstly, previous research indicated different levels of correlation between self-report and course test data. According to Pike, "the available evidence seems to indicate that the amount of content overlap between survey items and achievement tests is positively related to the magnitude of their correlation" (p.3). The second issue is the variation between the two types of measures. Objective or standardized tests tend to measure achievement very accurately over a narrow range of content, whereas, self-reports measure a much broader array of content with less accuracy. He suggests that if self-reports are to be used as a means of self-assessment instead of objective tests, "method-specific differences should be identified and accounted for statistically" (p.4).

These studies examining self-report data as a form of self-assessment, reinforce what Boud and Falchikov (1989) highlighted as a major problem with self-assessment research. They point out that commonly, student ratings of performance are derived differently from teacher scoring methods. Most often,

students are rating themselves using a survey format which can include a number of performance domains, whereas their evaluators are using traditional evaluation data such as essays and test marks. Boud and Falchikov (1989), state that even when students have been presented with the same valid measures or criteria for grading as the instructor (as they were in Anderson's study), they may not interpret the achievement standards in the same manner and may indeed arrive at different conclusions. This lack of a common understanding can have a significant impact on the validity of the self-assessment results.

The second technical issue requiring attention is reliability. In empirical terms, this refers to the stability, credibility and dependability of the assessment results. Reliability asks the question: Would this procedure produce similar results over time or on different occasions or for all students of a similar level; or across various assessors or raters? The most common measure of reliability used in the following studies given the qualitative nature of the assessment data appears to be inter-rater reliability or triangulation with other assessment results.

Ralph (1995), investigated the accuracy of self-reported data within a self-assessment framework from beginning teacher in relation to their

professional performance in their first year of teaching. The self-report data was triangulated with two other formal assessments of their professional skills. The beginning teachers completed a survey answering questions about their teaching performance. These data were then compared by the researcher to their final teaching practicum evaluation and a professional evaluation conducted by their school principal or school board administrator. The overall findings indicated that for the most part (90 per cent), self-report results were consistent with both the practicum as well as the professional evaluations. Ralph attributes these significant findings to a number of different causes which include: 1) The participants were assured of the confidentiality for their self-appraisals. 2) They were provided with specific training (although this training was not described), and were competent in with self-assessment procedures. 3) Both the participants and their supervisors valued and were committed to the self-assessment process.

Referring again to Higgins et al.'s (1994) study focusing on self assessment strategies at the first and second grade levels, individual student scores were viewed in comparison to their own progression of scores throughout the year, group project ratings, and teacher ratings. The findings highlighted two interesting results. Firstly, when student scores were

compared with teacher ratings, more than half the score were in complete agreement and when there were score differences, the teacher tended to score the students higher than they scored themselves (Higgins et al.). Secondly, although the peer groups tended to rate students lower on process criteria than they themselves did, there was 71% agreement on the product criteria ratings.

From the review of the literature, various important elements have arisen for further study. Firstly, the research has highlighted the need for explicitly stated learning expectations and criteria for self-assessment and the importance of ensuring a common interpretation of them; preferably achieved through student-teacher joint development of criteria and rubrics for performance interpretation. In addition to student and teacher cooperation, Anderson's findings suggested that exemplars be shared with students as a means of promoting a common understanding by further illustrating criteria and levels of achievement. Secondly, a number of the studies reviewed indicate the importance of training individuals to assess themselves competently if self-assessment data is to be valid and reliable. Thirdly, along with training, students need to be presented with and to practice using exemplars to illustrate various levels of achievement thus helping them interpret their own work quality. Next, triangulation or inter-rater comparison appears to be at this

time, the most appropriate means of determining the reliability of self-assessment when the interpretation of standards is similar for all parties involved. A fifth interesting issue, as highlighted by Higgins, Harris and Kuehn (1994), occurred when students developed their own criteria. These criteria were chosen in terms of importance relevant to the students as opposed to curriculum guidelines. However, if self-assessment is to be used as a means for judging student learning in the classroom setting, its validity with regard to curriculum guidelines must be clearly established. Unless this is accomplished, both students and teachers may see little value in the process and thus impact both the validity and reliability of self-assessment data. Lastly, as Boud and Falchikov stated, there has been a lack of focus in self-assessment research with a paucity of study replication and building upon other's postulations and discoveries. With this in mind, especially considering that little research has been done at the elementary level, this study is exploratory in nature and builds on aspects of the self-assessment process presented by Higgins, et al. (1994).

## Concepts and Issues Relating to Present Study

This final section of the chapter presents the research questions that have arisen from the literature review, a definition of terms as they are used in this study and lastly, the proposed conceptual framework that was used to guide the self-assessment process.

### Research questions

As stated initially, this study documents the implementation of a proposed self-assessment process in grades six to eight in an attempt to produce pertinent, reliable data. The research questions are as follows:

- 1) What is the process in which teachers and students participate in order to implement a proposed self-assessment framework with built in checks for producing valid and reliable results for interpretation?
- 2) What are the factors impacting the validity and reliability of the results?

### Definition of terms

According to Spandel and Stiggins (1990), assessment is a general term which means taking a closer look at student achievement in order to uncover strengths and weaknesses. Popham (1995), defines it as a formal attempt to determine students' status with respect to educational variables of interest.

Linn and Gronlund (1995), describe assessment as an integrated process incorporating a variety of procedures used to gain information about student performance and the formation of judgments or interpretations based on this information. Once intended outcomes have been assessed, the gathered data can be used for a variety of purposes including reporting, or improvement of learning and instruction. According to their definition, evaluation differs from assessment in that it takes the process one step beyond judgment to decision.

For the purposes of this study, assessment is defined as the process of identifying learners' strengths and weaknesses with respect to specified expectations. It is a multi-step process which involves: 1) Intention: statement of the purpose of assessment i.e. formative or summative, and identification of desired learning expectations, 2) Measurement: performance of a measurement task which will provide data about the learner's achievement in relation to the expectations and lastly, 3) Judgment: interpretation of the data using criteria which relate achievement to desired expectations. The proposed self-assessment process is broken down into each of the three steps and indicates the actions that are to be undertaken within each step. This is further discussed in the methodology section.

Self-assessment then, is the process which empowers the learner to be able to identify and understand the learning expectations, perform a measurement task, and interpret the data gathered by the task using criteria, rubrics and exemplars.

Expectations are defined by the Ontario Ministry of Education and Training (1997), as “the knowledge and skills that students are expected to develop and demonstrate in their class work, on tests and in various other activities on which their achievement is assessed.” (p. 48). Learning expectations are outlined for each grade from grades 1 through 8.

Criteria identify and describe the important aspects of a learning task and are used when making judgments about student performance of said task (Wiggins, 1993).

Rubrics are a common interpretation tool used for assessment in the elementary classroom. Spandel (1994), defines a rubric as a tool which provides criteria that describe student achievement at various levels of mastery. Rubrics are traditionally used by teachers as a judgment tool. In this study teachers and students participate together to choose criteria and create rubrics that the students use to judge their own work.

Lastly, the review of the literature indicated the importance of exemplars in order to illustrate various levels of achievement as well as to assist in the development of a common understanding of the performance criteria. Exemplars can be defined concrete examples of a performed measurement task which illustrate the various identified levels of achievement. For example, if a skill was to be assessed as novice, intermediate or advanced, exemplars would illustrate exactly what level of performance would be necessary to achieve each of the three rankings.

### Self-Assessment Conceptual Framework

The literature indicates that validity and reliability have often been compromised when self-assessment is used by learners to assess performance. In the studies examined, researchers have identified a number of factors which impact the usefulness of self-assessment and suggest a number of elements as being necessary to the self-assessment process in order to assure valid and reliable interpretation data. The proposed conceptual framework has been developed by the researcher to incorporate these elements highlighted by the literature review and other components considered of importance by the researcher within the self-assessment process. Gipps (1995), comments that assessment elements contributing to validity and reliability often cannot and should not be regarded separately. With this in mind, the researcher has identified and included elements that act jointly to produce data that are both valid and reliable. The framework components are as follows: a) the self-assessment process must be taught to participants; b) assessment takes place in a formative context; c) clear performance expectations and criteria must be evident and known to both students and teachers before the learning task is attempted; d) chosen learning expectations must relate to curriculum guidelines in order to be relevant to the classroom setting; e) students and teachers must

cooperatively develop rubrics which outline the selected performance criteria; f) outside expertise in the form of consultants, in addition to teaching peers must be consulted about rubric development in order to ensure that criteria descriptions closely relate to the to the learning expectation and clearly delineate between the identified achievement levels; g) rubrics are to be revised accordingly; h) students must be presented with exemplars which illustrate the different levels of performance defined by rubrics; i) students must practice with rubrics and exemplars in order to become familiar with the criteria and the self-assessment process; j) teachers must ensure that a common understanding of expectations, performance task and criteria for achievement interpretation has been reached before interpretation of performance takes place.

Ralph's (1995) study suggests that one of the aspects that contributed to valid and reliable assessment results was participant training. This training not only introduced participants to the self-assessment process but according to Ralph, also increased the participants commitment to the process. For the present study teacher training was necessary as the process was unfamiliar to the participating teachers. Training must be conducted in order to increase the teachers' comfort level with the process as well as to strengthen both the

assessment validity and reliability by demonstrating their importance in the assessment process. Common sense suggests that this framework element can be eliminated once teachers are familiar with the process and able to implement it with precision.

As indicated by the literature review, the self-assessment process calls for explicitly stated learning expectations and judgment criteria, and underlines the importance of ensuring a common understanding of these elements amongst students and teachers. Without clearly defined learning expectations and criteria, assessment validity will be compromised. In order to guarantee inter-rater reliability, a common understanding of these elements is vital. Studies reviewed showed that common understanding of criteria was enhanced when student and teachers worked together to develop assessment criteria (Higgins, et al., 1994; Mckinney, et al., 1995). Higgins et al. worked with students to reinforce their understanding of the chosen assessment criteria and checked with students to verify that this understanding was commonly held. They attributed a 71% agreement rate between students and the teacher on product criteria to this.

The use of exemplars in order to illustrate the achievement levels which are described by the student-teacher developed rubrics is a second important

factor relating to developing a common understanding. As mentioned previously, Anderson states that besides clearly stated learning objectives and criteria, students must be presented with performance exemplars to foster a common understanding between students and teachers of the assessment criteria (1993). Joint rubric development along with the presentation of exemplars are important elements which contribute to reliability of assessment data.

Four elements not previously emphasized by the reviewed studies have been included by the researcher in the conceptual framework: ensuring that learning expectations are relevant to curriculum guidelines; setting a formative context for assessment; allowing for student practice with rubrics and exemplars; and lastly, using outside expertise as well as teaching peer feedback to strengthen the validity of the assessment rubrics through revision.

As stated in the literature review, a paucity of research exists which examines the use of self-assessment in the school classroom. In order for the self-assessment process to become a useful means of alternative assessment in the school classroom it must be manageable for use in that context and perhaps more importantly, it must relate to curriculum guidelines. Unless these two

qualifications are met, there is little hope that the self-assessment process will be incorporated into the existing classroom assessment scheme.

Stiggins (1987), suggests that self-assessment is most effective when the situation is not a high stakes one where students have something to gain from artificially inflating their scores. With respect to this, the self-assessment framework is used in this study for formative purposes, with students assessing their writing in language arts as part the writing and editing process.

Practice with criteria rubrics and exemplars is a another element which is designed to increase assessment data reliability. Practice is included in the conceptual framework for two reasons. Firstly, practice time can be used by teachers as a forum for discussion about expectations and criteria thus contributing to a common understanding. Secondly, practice facilitates student familiarity with the process a factor which is indirectly implied in the literature review. When Higgins, et al. (1994), conducted their study, they allowed time for their students to become comfortable with the measurement concepts and criteria before self-assessment data was collected. This researcher included practice as a framework element with the expectation that it will promote a common understanding of criteria as well as increase students' comfort level with the process thus strengthening the reliability of the assessment.

Lastly, according to educational experts (Wiggins, 1993; Stiggins & Conklin, 1992), teachers often receive inadequate training in the areas of measurement and assessment. Involving outside expertise in the process along with providing the participating teachers with the opportunity for feedback will both inform and strengthen the process.

Obviously, some of the elements of the framework must be implemented in a chronological fashion, while others, once implemented should be revisited by students and teachers throughout the process. The procedure section in the following chapter illustrates how the elements are implemented chronologically during the three steps of the assessment process. The elements of the process are also described and include: a) seeking outside expertise to strengthen validity, b) rubric revision, c) student practice with exemplars and rubrics and d) ensuring a common understanding of the expectation, criteria and the overall process. The framework's flexibility allow teachers to decide the level of emphasis on each element. (See figure 1, p.37).

This conceptual framework is designed to include important elements of the self-assessment process yet remain flexible to allow implementation variation within the process. As this was the first time the participating teachers had attempted to use student self-assessment, an implementation plan

was followed which broke down the process into three steps: intention, measurement and judgment. Each of the three steps was further broken down into objectives to be accomplished within each step. The goal of the researcher intervention and introduction of the framework was to empower teachers to follow the proposed self-assessment process within their classrooms while at the same time providing them with a structure to ease the implementation process. Using language arts expectations, students and teachers were to identify important criteria and develop rubrics which described levels of mastery which were then to be used as an interpretation tool to rate the measurement data. Both the study methodology and the intervention are discussed in detail in the following chapter.

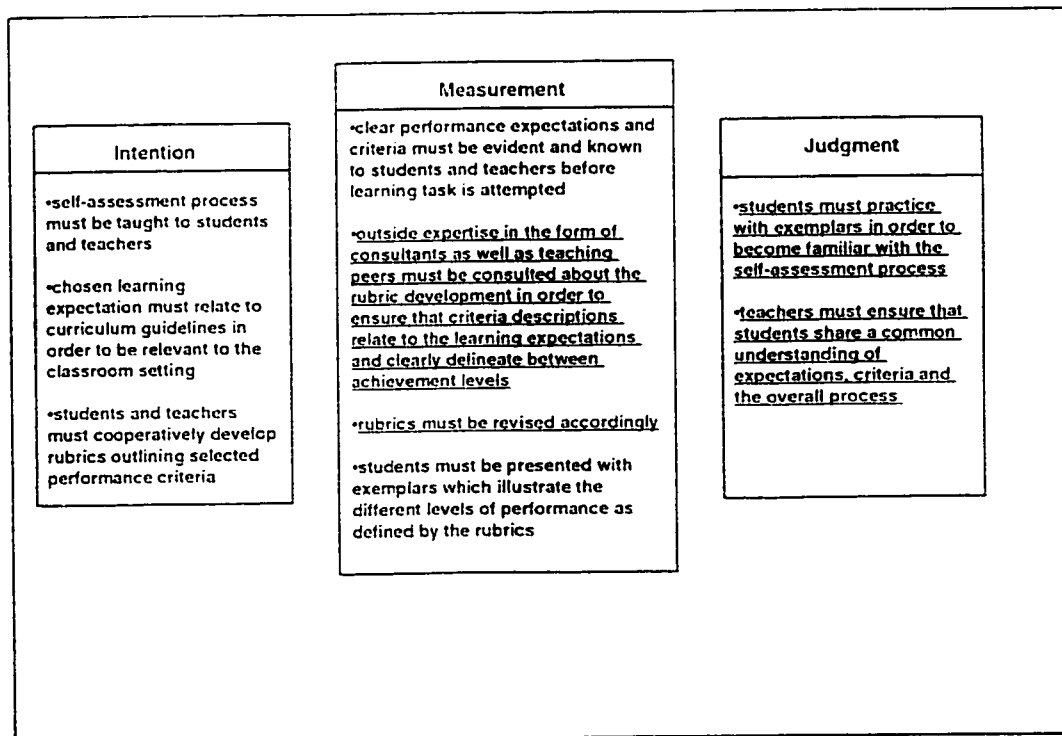


Figure 1: Student self-assessment conceptual framework. Chronologically important elements are presented as they occur in the assessment process. Elements that are considered ongoing and must be revisited throughout the process are underlined.

## Chapter 3

### Method

The main focus of this study consisted of examining the implementation process for formative student self-assessment in a classroom setting in an eastern Ontario School Board during the 1997-1998 school year. As stated, the purpose was to document the assessment process in which students and teachers co-participated in order to produce both valid and reliable results. Details regarding the participants, instrumentation and the procedure are provided in the following sections.

#### Participants

##### Teachers and students

Six grade 6, 7 and 8 teachers from three schools in Eastern Ontario and seventy four of their students participated in the study during the 1997-1998 school year. As this was an implementation study and required teacher participants who were interested in learning about self-assessment, the selection process was purposive rather than random. After receiving permission to conduct the study from the board superintendent, the researcher

solicited teachers volunteers with the help of the school board's junior/intermediate language consultant. The criteria for selection were simply an interest in alternative assessment strategies and a willingness to involve students in the assessment process. Ten teachers were invited to participate. Of the original ten, three agreed to participate. The remaining three participants were interested peers of these three.

Of the six teachers involved in the study, all had experience using some form of alternative assessment, for example; student lead conferencing, portfolios, peer evaluations. All had used rubrics to assess students' work and had designed rubrics for student assessment purposes. None had involved students in the rubric development process, nor had followed the self-assessment process described in this study. (See Table 1 for a description of their teaching experience, grade level presently teaching and alternative assessment background, p.40). Each of the teachers had attended workshops about the new Ontario Language Curriculum in 1997 and was using it in the classrooms for teaching and assessment purposes.

Although all of the students in each class participated in the self-assessment process; in total, 74 students received permission to share their self-assessment results with the researcher. Those who did not receive permission

participated in the process but their self-assessment data was withheld from the researcher. The six classes consisted of mixed socio-economic and family

Table 1 Teacher Participants

<u>Teacher</u>	<u>Years Teaching Experience</u>	<u>Grade Levels Currently Teaching</u>	<u>Alternative Assessment Experience</u>
Kate	13	8	portfolios, student led conferencing
Diane	10	7	portfolios, student conferencing
Ally	7	6	portfolios, peer conferencing
Peter	2	7/8	phys. ed. self-evaluation forms, peer conferencing, portfolios
Amy	1	7/8	phys. ed. self-evaluation forms, peer assessment
Lauren	1	7	portfolios

Note. pseudonyms have been used to ensure anonymity

backgrounds. Approximately 30% of the overall student population had been identified as needing language resource help or as learning disabled.

As the teachers were to work together, each class assessed a similar set of expectations as designated by the provincial curriculum (appropriate for each grade level). Writing was the complex skill that was focused on and the specific expectations that were the object of the assessment were chosen by the participating teachers in the first portion of the study. Learning expectations

for all areas of language arts are outlined in *The Ontario Language Curriculum: Grades 1-8 (1997)*.

### Consultants

Two consultants participated in the study: the school board language arts consultant and a representative of the Ministry of Education and Training. The language arts consultant from the school board participated in the four team meetings. Her school board responsibilities include liberal arts education for kindergarten to grade 8. She has attended numerous workshops on the new Language Arts curriculum and offered expert opinion on the topics of rubric development and exemplars. A consultant from the Ministry of Education and Training was the second expert to provide guidance to the teachers within the process. Her current role for the Ministry is to develop exemplars to accompany the new Ontario curriculum for language arts. Her expert opinion was sought regarding the validity of the exemplars chosen by the teaching team as well as the rubric development.

### Researcher background

A brief description of the researcher is provided as she played an important role in the implementation of the intervention. She is an junior/intermediate teacher with five years teaching experience. As her

teaching career progressed she became cognizant of the fact that her university education had ill prepared her to conduct quality student evaluation and assessment. As a teacher, she felt that traditional evaluation practices did not always serve the learner as well as they could and felt this type of evaluation often discouraged rather than informed many learners. She returned to university with the wish to learn more about assessment and to investigate methods that provide valid, reliable results and are designed rather than to just monitor or audit student performance, enable learners to improve it.

The researchers main role in the study was one of guidance, specifically with the goal of empowering teachers to implement the process in their classrooms. Secondly, she provided expert opinion with regard to external technical checks. She was also responsible for collecting descriptive data to document the process. Each of these roles are further described in the procedure section.

### Instrumentation

In order to gather and analyze data regarding the implementation of the proposed framework for self-assessment within the classroom, two instruments have been developed by the researcher: a double entry journal, and an

interview guide. A third means of gathering information was the researcher's field notes and observations of the meetings.

A double entry journal (see Appendix A) was set up with two columns for each entry. One column was used to document purposes while the second asked the journal writer to reflect/critique the process that was recorded. Specific entry titles provided by the researcher for the teachers are as follows:

- 1) Discussion of writing expectation and criteria with students, ensuring validity.
- 2) Development of 1st draft assessment rubric.
- 3) Discussion of exemplars and changes to rubrics.
- 4) Practice with the rubric.
- 5) Final draft of the rubric.
- 6) Use of rubric for student self-assessment.
- 7) Peer assessment of work.

This allowed teachers to chronicle the self-assessment process as they implemented it with their students, as well as providing them with an opportunity to critically reflect on the process as each step was completed. These journals provided the researcher with a means of documenting internal validity and reliability checks as well as gathering teachers' critical reflections regarding the overall process.

Interview guides (see Appendix A) were developed for the participating teachers in order to capture qualitative data concerning the overall self-assessment process and its technical qualities as it was applied within a

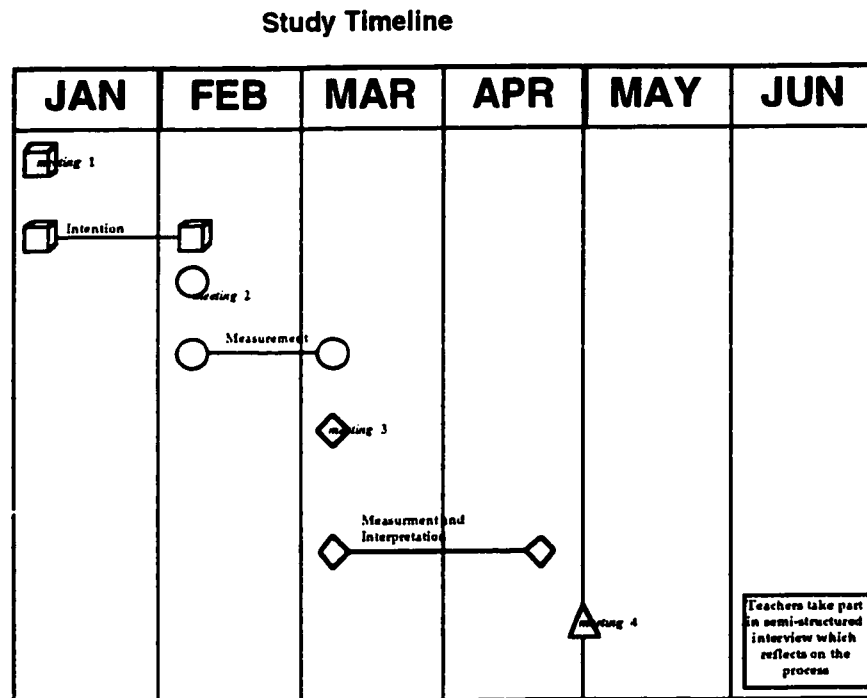
classroom context. The questions were framed in part by the self assessment framework in order to uncover data relating to the validity and reliability of the process. A second goal of the interview was to capture factors that impacted the process. The interview guide was semi-structured to allow for emergent responses. It allowed the researcher to probe beyond the information contained in the journals as well as to gain the teachers' professional observations about the students' participation in the process. Interviews were taped with the teachers' permission.

### Procedure

Ideally, this study would investigate an existing self-assessment scheme and its technical qualities thus reducing the intervention role of the researcher. However, because the teachers were implementing a framework that is researcher guided, the researcher occupied an important role in the process. In an attempt to reduce the possibility of researcher bias, the researcher provided teachers with a variety of resources relating to the self-assessment process and rubric development in order to help teachers internalize the process before classroom implementation. This empowerment was further developed through monthly full team meetings which are described in detail in this

section. She also brought outside expertise in the form of the school board's language arts consultant as well as a Ministry of Education and Training Consultant to assist with external validity and reliability checks. For these particular reasons, the researcher maintained the role of facilitator and data collector.

The study took place during a period of six months. At the beginning of the months of January, February, March and May the researcher conducted a meeting which was attended by the participating teachers as well as the school board's junior/intermediate consultant. A total of four meetings were conducted during the period of the study. A study time line is illustrated in figure 2. The external (meetings) and internal (in class) procedures are described as they relate to intention, measurement and interpretation in the following section.



**Figure 2:** Overview of study timeline

As previously mentioned, because the participating teachers were unfamiliar with the process, the researcher provided them with an implementation plan. This plan is discussed in the following section and is illustrated in figure 3 (See p. 50).

**Step One: Intention:** The researcher introduced teachers to the goals and outline of the study, the framework that they were to use as a guide for implementing the self-assessment process with their students, as well as the data collection materials. A mini-workshop was conducted relating to

measurement concepts and principles to ensure that teachers shared a common understanding of the important concepts relating to the study. The board consultant assisted the researcher by helping the teachers to choose a combination of writing expectations that were appropriate for each grade level.

Before the second meeting, the teachers were requested to introduce the students to the project and were to attempt to ensure that their students shared a common understanding of the expectations and performance criteria. They were requested to come to the second meeting with a draft rubric and samples of student writing which might be used as exemplars to illustrate the four levels of achievement according to the ministry guidelines.

### Step Two: Measurement

Teachers met with the board consultant and researcher to discuss their initial rubrics and choose exemplars which illustrated the four levels of achievement. Exemplars were to be chosen from a pool of writing samples gathered by the teachers and the researcher. The team was to work together to evaluate each class' rubric and suggest improvement to improve validity by ensuring that the criteria related to the learning expectation and reliability by suggesting improvements in language clarity.

Rubrics were forwarded to a Ministry of Education and Training consultant for feedback regarding validity of the rubric criteria and descriptions. Comments were also solicited on the choice of exemplars for each of the three grades. These comments were to be communicated to the teachers before the third meeting.

Within their classrooms, teachers were expected to implement the second block of the framework. During this month, exemplars were to be presented to the students as well as a means of illustrating the four achievement levels appropriate for each grade thus further promoting a common understanding of the criteria and increasing reliability. A second reliability check asked the teachers to allow students time to practice and familiarize themselves with the use of the rubrics.

### Step Three: Measurement and Judgment

Teachers brought rubrics meeting for a final critical analysis. Feedback from the Ministry of Education and Training consultant was discussed. Any final discussion regarding rubric development took place at this meeting. Measurement tasks were discussed and each teacher decided with the help of the team, on an appropriate writing task.

In the classroom, students were to complete the first draft of an assigned writing piece and to interpret it in relation to the specified learning expectations by using the developed rubric. Once a student had used the rubric to interpret the data, two other interpretations were collected from the teacher and a student peer in order to determine the reliability of the results.

#### Step Four: Debriefing

This last meeting was a debriefing session to discuss the success of the process. Teachers shared the final month's experiences as well as their opinions of the process. Student assessment data, final copies of the rubrics and the double entry journals were collected by the researcher at this time.

Semi-structured interviews were scheduled by the researcher with the teachers approximately month after the final meeting in order to accommodate the teachers' busy schedules. (The interview guide is presented in Appendix A).

PROCESS	INTENTION		MEASUREMENT		MEASUREMENT AND INTERPRETATION	
	Meeting One	Meeting Two	Meeting Three	Meeting Four		
<b>External (Meetings)</b>	<ul style="list-style-type: none"> <li>introduce framework to teachers, indicate that self-assessment is to be formative.</li> <li>establish learning &amp; assessment goals.</li> <li>identify expectations that students and teachers will be using.</li> <li>workshop teachers on assessment &amp; measurement concepts.</li> <li>assure common understanding of expectation and process (r)</li> </ul>	<ul style="list-style-type: none"> <li>researcher, teachers and board consultant evaluate rubrics in relation to expectations. Suggestions for improvement are made (v)</li> <li>exemplars are chosen to illustrate the four levels of achievement for each grade level.(r)</li> <li>rubrics and exemplars are sent to MET consultant to check adequacy and relevancy of criteria and exemplars in relation to expectation. (v)</li> </ul>	<ul style="list-style-type: none"> <li>teachers receive feedback from MET consultant (v)</li> <li>final discussion as to improvements to be made to the rubrics. Same as meeting two.(r&amp;v)</li> <li>discussion relating to appropriate writing task and teachers each decide upon a task for their students (v) .</li> <li>Students will use rubrics to interpret their first draft of the assignment</li> </ul>	<ul style="list-style-type: none"> <li>teachers bring self, peer and teacher ratings to the researcher for calculation of agreement percentages. (r)</li> <li>final meeting used as a debriefing session as well as a chance for teachers to share their opinions regarding the final steps of the process</li> </ul>		
<b>Internal (Classroom Process Occurred after each team meeting)</b>	<ul style="list-style-type: none"> <li>students are introduced to the goals of the study and their role.</li> <li>students are presented with the chosen expectations for learning.</li> <li>teachers and students work together to identify important criteria within the expectations. (v)</li> <li>teachers attempt to ensure that students share a common understanding of the expectations as well as the identified criteria. (r)</li> <li>students &amp; teachers work together to develop 1st draft of performance rubric. Whole class participation to ensure common understanding (r) and understanding of rubric's relevance to the expectation (v)</li> </ul>	<ul style="list-style-type: none"> <li>students and teachers use meeting suggestions to improve rubrics in terms of relevancy (v) and common interpretation (r)</li> <li>students are familiarized with the exemplars and practice using rubrics on the exemplars and other pieces of work (r)</li> </ul>	<ul style="list-style-type: none"> <li>final changes are made to the rubrics (v&amp;r)</li> <li>students create draft piece of writing.</li> <li>students use rubric to interpret their performance regarding their first draft.</li> <li>both the teacher and a peer use the rubric to assess the same first draft.(r)</li> </ul>	<ul style="list-style-type: none"> <li>students have completed their participation in the study</li> </ul>		

**Legend: (R) = reliability checks, (V) = validity checks**

**Figure 3:** Implementation framework outlining external and internal aspects of the process

## Analysis

After completing the interviews, the audio tapes were transcribed verbatim. As well, double entry journals, field notes and observations were combined with the interview data. Codes were applied to identify the external and internal validity checks as well as commonalties and variations with these checks between classes in the implementation process. Specifically, steps that were taken that related to validity and reliability were noted and categorized. At the same time themes and salient points identifying factors affecting the validity and reliability of the process were highlighted.

Once the student self-assessment data was collected, the researcher determined inter-rater percentage agreement between student and teacher, as well as student and peer scores in order to determine the reliability of the students' self-assessment scores.

Results are presented in the following chapter. Meetings are documented and external technical checks are highlighted. In order to incorporate the variation between classroom implementations, a synopsis of the process as it occurred in each class is presented.

In response to the second research question, the factors influencing the technical aspects of the self-assessment process as they emerged from the

collected data are presented. A checklist matrix is used to highlight the major factors as they were identified by both the teachers and the researcher and indicates whether the effect was a negative or positive influence on the technical qualities of the process.

## Chapter 4

### Results

As the study is composed of two research questions, this section presenting the results of the study is divided into two sections. The first question asked what is the implementation process within a classroom setting of the self-assessment framework with built in checks to produce valid and reliable results? The second asked what factors arose in the classroom implementation process impacting the validity and reliability of the results? The documentation section provides a description of the actual external meeting process as well as synopses of the internal process as it occurred in each classroom. It also highlights the results of the validity and reliability checks as they occurred. It is then followed by a section presenting the elements impacting the process that arose from the double entry journals and the interviews as well as the researcher's personal field notes.

#### The Self-Assessment Process

In order to illustrate the self -assessment process as it was implemented, this section is presented in three parts. The first section describes the external process as it occurred. Secondly, the collected self-assessment data is analyzed

for inter-rater agreement. The last part presents the reader with a synopsis of the internal process as it occurred in each classroom.

### External Process and Technical checks

Meetings were held for essentially two purposes: to help facilitate the implementation process for the teachers and their students and to provide a means for establishing external as well as internal checks for validity and reliability. As stated previously in the procedure section, teachers met as a group with the researcher and the board consultant a total of four times. A description of the external process as well as technical quality checks is presented in this section.

At the first meeting, the teachers were given a study outline describing the timeline for framework implementation as well as the steps that would be taken to ensure validity and reliability. As stated in the procedure, a mini workshop was presented by the researcher to the teachers and focused on measurement concepts, specifically validity and reliability. Rubrics were discussed as a judgment tool and rubric development was reviewed with technical qualities in mind. A binder was distributed to the participants which outlined the framework for implementation, contained the double entry journals as well as a resource package which contained a number of

professional articles relating to rubric design and self-assessment. For the purposes of the study, student assessment was presented to the teachers as a three step process: intention, measurement and judgment/interpretation. The researcher reviewed the framework indicating the external and internal steps as well as the technical checks that would occur during the student self-assessment process.

One of the main goals of this meeting was to decide upon the first of the three steps: intention of the assessment. With the help of the board consultant, the teachers decided to combine two of the expectations from the Ministry curriculum guidelines for language arts relating to paragraph construction. These were chosen because of the global applicability of these expectations to writing in a variety of subject areas.

The chosen expectations are outlined as follows:

Grade 6: Students will organize information to develop a central idea using well-linked and well-developed paragraphs as well as a variety of sentence types.

Grade 7: Students will organize information to develop a central idea, using well-linked and well-developed paragraphs as well as use a variety of sentence types and lengths.

Grade 8: Students will organize information and ideas creatively as well as logically, using paragraph structures appropriate for their purpose (e.g. paragraph structure to develop a comparison or establish a cause-effect relationship and use a wide variety of sentence types and sentence structures.

At the end of the first meeting, teachers were asked to implement the elements outlined in the Intention section of the implementation framework for meeting two.

Teachers arrived at the second meeting with the first draft of the rubric that they had developed with their students. At this point in time, the rubrics varied widely. For example two of the rubrics were jumbled lists of criteria lacking a thematic focus. Two other classes of students decided with their teachers that they would not include descriptions of level four performance in their rubric. These students were unwilling to include level 4 targets they felt that none of them could achieve. The teachers had ensured that the whole class was in agreement with this decision. Students decided that they would include level four when they were more capable writers.

The teachers, the board consultant and the researcher reviewed each of the rubrics with a discussion focusing on validity of the criteria as they related to the learning expectation and the descriptions as to how well they

represented the provincial achievement levels. Suggestions for improvement were noted by the teachers to take back to their students.

In order to assess validity, an exercise was conducted where the teachers identified the criteria within the expectation. The class rubrics were then compared to ensure that the criteria the classes chose met with the criteria identified by the meeting group. Criteria that had been included in the rubric but did not meet with the expectation were highlighted for change. To ensure reliability, the language of the rubric was then evaluated for clarity and ease of understanding. Overly general terms such as "good" or "excellent" were noted in order that students and teachers could devise more explicit phrasing to promote a common understanding. It was decided quantitative phrasing would be suggested where possible to make the rubrics easier for the students to understand.

Prior to this meeting the teachers and the researcher had collected a large pool of a variety of student writing samples from which exemplars were chosen. Pieces were anonymous and steps were taken to ensure that students would not be presented with their own or their classmates work. With the help of the board consultant, four pieces of writing were chosen for each grade level in order to illustrate each achievement level thus promoting a common

understanding of each level amongst the students and strengthening reliability. The draft rubrics and exemplars were then forwarded to a ministry consultant for feedback regarding the validity of each. The consultant returned the rubrics with improvement suggestions to the researcher who then forwarded them to the teachers. Improvement suggestions were similar to the ones decided upon at the second meeting and related to clarity of language. Also, criteria such as spelling or vocabulary were identified as not relating to the chosen expectations and therefore should be removed. The chosen exemplars were approved (See appendix D for each grade level's four exemplars).

During the third meeting, each teacher presented his or her class' edited rubrics to the group. Suggestions were made to improve validity and reliability once again. At this point the rubrics' technical quality varied considerably between classes. For example four of the six teachers decided to disregard the suggestion to remove criteria describing language conventions because they had discovered that their students' work would suffer in this area if they weren't included in the assessment, thus having an impact of the validity of the assessment. Although this represented a significant gap in students understanding of the expectation as it related to writing tasks, after a lengthy discussion, these teachers decided it was easier management-wise to leave

things as they stood. Suggested final changes to the rubrics, if any, were returned to the students to incorporate into their final valid edition of the rubric, although most teachers agreed that their students were unwilling to make further changes as they felt they had a finished product by this point. Teachers promised that they would discuss this issue with their students but would let them decide.

According to the teachers at this point in time, the majority of the students were comfortable using the rubrics and demonstrated a reasonable understanding of how the exemplars represented each achievement level. A discussion was led by the researcher to determine appropriate measurement tasks for the study data. The teachers decided to ask their students to write a multi-paragraph expository piece about a topic specified by the teachers. For example, one assignment was to describe the perfect dining experience.

At the final meeting each teacher presented the researcher with his or her double entry journals, a copy of the final rubric and the data collected relating to the student self-assessments (See Appendix B for copies of the final rubrics). Self-assessment data included a self rating for each student as well as a peer and a teacher rating. These ratings were compared by the researcher in order to determine inter-rater reliability and is presented in the following section. The

researcher's main goal for this meeting was to provide teachers with a chance to share ideas and reflections about the process. Although there were concerns expressed about the students' ability to assess such a complex performance task, all teachers concurred that it was a valuable process and stated that they would reuse this process in the future.

### Analysis of self assessment data

As mentioned in the literature review, the most common means of determining data reliability with qualitative assessment is to establish the inter-rater reliability. In order to triangulate the student self-assessment data, teacher and peer assessments were done using the rubric as a judgment tool for each individual's work. The following is a presentation of the percentage agreement within the group at large as well as within individual classes is presented here.

A total of three hundred and forty nine self assessment scores were collected along with a similar number of teacher and peer scores. The following tables present the inter-rater agreement in percent.

**Table 2: Overall inter-rater agreement**

student self-assessment & teacher agreement	student self-assessment & peer agreement
66%	54%

**Table 3: Inter-rater agreement within classes**

raters	Kate	Dawn	Ally	Peter	Amy	Lauren
student/teacher	54%	77%	65%	91%	53%	42%
student/peer	54%	51%	50%	64%	55%	35%

**Internal Process**

The framework and methodology provided in the meetings were to provide structure, yet empower each teacher to complete each step of the process as he or she saw fit. It also served to incorporate technical checks within the classroom. In order to describe the self-assessment within each classroom, a summary of the double entry journal and interview data which documents the internal process for each classroom is presented. A synopsis of each teacher’s classroom implementation is presented in table form in Appendix C.

Kate: grade 7/8

student/teacher agreement: 54%

student/peer agreement: 54%

Kate introduced the expectations to her students and discussed the meaning of criteria and how it applied to this expectation. The whole class worked together to choose the criteria to be assessed and describe performance for levels 1-4. Once the rubrics were completed, they used them to judge the exemplars. Kate had hoped that this might prompt them to make changes in the rubrics to improve validity and reliability however, the students declared themselves content with the rubric as it was. Any attempts on Kate's part to discuss the rubric's quality resulted in very low participation and discipline problems. She decided to "leave things as they were." Due to lack of interest and time constraints, no further changes were made to the rubrics. Students practiced using the rubric on one or two pieces before providing the self-assess data for a draft piece.

Little was done to ensure validity beyond the initial activities involving the whole class. Kate described circulating and providing individual feedback to help students stay with the criteria described in the rubric when doing self and peer ratings. As mentioned, Kate suggested changes to the rubric but decided for management purposes to leave it in its first draft form. To ensure reliability, as the students worked with the exemplars and practiced with the rubrics Kate provided individual and large group feedback relating the descriptions to actual performance.

Diane: grade 7

student/teacher agreement: 77%      student/peer agreement: 51%

Diane presented her students with the expectation and the ministry achievement levels (found in the curriculum guide). Each of these was discussed with the whole class. They also discussed "criteria" and identified some of the criteria within the expectation. In small cooperative groups students designed a rubric for the expectation. Diane combined the small groups' products to produce a rubric for the whole class.

This initial rubric was very general and had not been divided into categories. Once they received the exemplars, Diane and her students grouped common factors together in order to identify specific criteria and performance descriptions. Once this was done, the students used the rubric to practice assessing a variety of pieces. After the third meeting, the students re-evaluated the rubric and made last minor changes. They then used the rubric to self-assess the first draft of a multi-paragraph piece.

Diane attempted to ensure that all students understood the expectations and criteria and how they related to the skill of writing through small and large group discussion. She found that some of her students did not understand some of the concepts that were to be assessed and spent extra time reteaching. When practicing with the rubrics, Diane had the students do peer reviews in pairs and justify their judgments to their peers. Difficulties were addressed in small and large group discussion. These latter steps were taken to ensure a common understanding of the assessment criteria.

Ally: grade 6

student/teacher agreement: 65%      student/peer agreement: 50%

Ally presented the expectation to her student and discussed what it meant and how it could be evaluated. She used the ministry achievement levels as a model for rubric development. Students worked as a class and were guided to develop a rubric that would assess the expectation. After the second meeting, the class separated the rubric into categories and rephrased some of the descriptions. They used the exemplars as a practice assessment to decide which piece would be assigned to each level. The class spent a lot of time discussing what level 3 quality work entailed. A final draft was completed after the third meeting and more self-assessment practice was done with reteaching where Ally deemed it necessary. Once students felt confident, they self-assessed a draft piece for the study.

To ensure validity, Ally constantly brought students back to rubric descriptions and discussed how they applied to the writing samples. Much reteaching of concepts was done to enable the students to understand what they were assessing. As for reliability, again, time was spent discussing rubrics and examples of performance in order to ensure that students shared a common understanding of each level of performance.

Peter: grade 7/8

student/teacher agreement: 91%      student/peer agreement: 64%

Peter explained to his students that they would be developing a rubric to assess the chosen expectation. He and his students looked at sample rubrics and he presented how he would develop a rubric for assessment. He broke students into small groups to identify and describe criteria with little success. He then worked with the whole class to identify and describe criteria for a rubric. After much discussion, the students decided to include only three levels in the rubric.

Once the initial rubric was developed the students used it to assess the exemplars, Peter was pleased to note that the majority of the students assigned the correct level to each piece. Students also decided upon changes that should be made to the rubric. Students practiced using the rubric on a number of draft pieces and Peter encouraged them to use it as a checklist when creating a piece. For the most part students assessed in pairs to practice justifying their interpretation thus strengthening reliability of the assessments.

When collecting self-assessment data, Peter modified the tool for students who were having difficulties. These students were allowed to choose to criteria that they felt comfortable assessing. The final peer assessments were done in pairs in order to increase accountability and hopefully accuracy.

In order to strengthen validity, Peter spent a great deal of time with his students discussing the concepts that they were addressing with their rubric and how their own writing related to these concepts. He also conferenced with each student as they practiced the self-assessment process to identify areas where each student was having difficulty. Regarding reliability, Peter's motto was Practice, practice, practice. He and his students incorporated the process into

any writing exercise that they undertook and by the time that the data was collected, the students were very familiar with the interpretation tool.

Amy: grade 7/8

student/teacher agreement: 53%

student/peer agreement: 55%

Initially, Amy reviewed the assessment process of leveling performance from 1-4. The chosen writing expectation was then introduced to the class and the whole class worked together to identify the criteria that they wanted to assess within the expectation. In pairs the students developed performance descriptions for levels 1-3. The students decided, against Amy's wishes to not include level 4 because they felt it was unattainable at the present time. Amy collected the student descriptions and combined them to form a rubric for the whole class.

Next, they used the rubrics with the exemplars to become familiar with each of the levels and they edited the rubrics to make language clearer and more specific. The rubric was then used as a checklist to practice peer and self-assessing. With the final rubrics students assessed a first draft of a creative writing assignment and then a peer assessed it as well.

Amy spent time with students individually as well as with the whole class to help the students understand the criteria and the expectation in an attempt to make the process a valid one. Students worked together to evaluate exemplars and to do peer assessments in order to enable a common understanding.

Lauren: grade 7

student/teacher agreement: 42%

student/peer agreement: 35%

Lauren introduced the expectation to the class and they discussed the meaning of it. They then examined a sample rubric and used it to help define criteria. As a class, they identified the general criteria categories and in small groups designed a rubric for each of the categories. Lauren assimilated these rubrics to produce a general one for the whole class.

Lauren used the exemplars with her students to illustrate the various mastery levels and to fine tune their rubrics. The rubrics were then used to assess and discuss different types of writing. No changes were made after the final meeting and the students used the rubric as a checklist to self-assess a piece of their writing. They were each given an anonymous peer's work to assess. These assessments were collected for the researcher.

While creating and revising the rubric, Lauren discussed with her students what each of the criteria meant in an attempt to enable them to make valid self-assessments. Class discussions were led when evaluating the exemplars as to why each exemplar represented a particular level but Lauren felt that a common understanding was not achieved as is evidenced by the inter-rater percentage agreement for her class.

### Similarities and Differences Between Class Groups

As each of the above summaries describes, there are commonalities and variations between the classes as to how the self-assessment process occurred. An interesting issue relating to this is the difference between classes with regard to inter-rater agreement. Student/teacher agreement ranges from a low of 42% to a high of 91% indicating a significant difference between classes with regard to technical quality of the self-assessments. These important instances are discussed here and Peter's results are used as a contrasting example.

The effective integration of a self-assessment process within a classroom assessment scheme empowers its users, incorporates user flexibility while at the same time producing results that are of value to the learner and educator. At first glance when one reviews each teacher's synopsis of the process, it appears that each class followed a similar process in implementing each of the various elements of the conceptual framework. Although each teacher incorporated the elements as directed, interestingly, the variations between classes occurred in relation to the degree of implementation of each element rather than element inclusion.

Although variables beyond the ones focused on in this study might account for the variation in rater agreement between classes, when one

examines the process descriptions, one difference between Peter and the other teachers is continually highlighted: teacher level of commitment to the process.

In short, Peter implemented each of the elements of the conceptual framework as did the other teachers. The difference lies in the level to which each element was implemented. His enthusiasm for the project and his high level of commitment was indicated by his descriptions of the classroom process. Not only did he provide his students with reteaching for difficult concepts and much opportunity for practice, as did some of the other teachers, he also assured that the self-assessment process became an integral part of the writing process for his students. The rubrics became a tool that his students referred to as a regular part of the editing process rather than something they used only when they were providing study data. His enthusiasm and commitment to the process were evidenced by his journal entries and his class' assessment results.

### Factors Impacting Technical Quality

Essentially, the goal of the study was to implement a framework that would provide both valid and reliable results for students and teachers. As

indicated in the process documentation, a number of steps were built into the process in an attempt to ensure the technical quality of the results.

As indicated in both the procedure and results sections, external validity and reliability checks were built into the meetings and involved teachers, the board consultant, the researcher and the Ministry consultant in the decision making process. The researcher's goal for the external process was to empower teachers to implement the self-assessment process in their own classrooms. Teachers were not forced to follow the framework rigidly, instead, it was left to the teachers discretion to consider the context of their classrooms and implement the framework as they saw fit. As demonstrated by the process documentation results, there were variations between classes in how each teacher and his or her students dealt with the technical aspects of the process.

When reviewing the data, a number of interesting, common factors arose within the internal process that impacted the technical quality of the self-assessment results. In the meetings, journals and interviews, teachers repeatedly identified factors associated with rubric development and practice as having the largest impact on the process. Other common factors include ensuring a common understanding of the expectations and the rubrics, and

student comfort level or student ownership of the process. These factors are discussed in this section.

### Achieving a Common Understanding Across Ability Levels

Gipps (1994), has stated that the students' understanding of learning criteria is a key component of formative self-assessment. Ensuring a common understanding of all of the criteria was a difficult task for some of the teachers. They felt that the varying levels of students language academic abilities had an impact on the children's ability to understand some of the paragraph writing concepts.

I've got several students in my class who are IPRC'd and a couple of ESL students and they had a very difficult time understanding the concepts. I was thinking I would have to modify the rubric for them. (Amy)

I found that the kids who were excellent writers could identify a well developed idea because they know they need to give examples. They know that they need evidence to back things up...but other kids would just sort of say...Well it's longer, the sentences are longer, therefore it's more well developed. (Ally)

Some of the teachers felt that the chosen expectation was too conceptually difficult for the students to develop criteria and descriptions that would produce both valid and reliable results.

Next time I'd try to level the playing field a bit by ensuring that whatever is on the rubric is completely 100% understood by every student. So in other words, some of them just didn't get what "evidence" meant while others did. Well I shouldn't have

been using a rubric with that on it. In other words, I would either have to teach these concepts more thoroughly so that anyone can look at a piece and say Now that's evidence or that's a topic sentence, or just not include it. (Peter)

The next time I do this, I'll focus on something less technical, less narrow. I'd choose something more global like choosing criteria for assessing a project. (Lauren)

One of the teachers felt that students had a great deal of difficulty limiting their focus to the expectations and suggested that perhaps a different approach to identifying specified achievement criteria might be easier.

Assessing specific outcomes may be too technical for students to grasp. Ironically, I think assessing a more global complex skill like Amy did with a second rubric focusing on science labs would probably be easier for the students. Even though this was a complex skill, because this was so specific the kids found it hard to break down. (Peter)

Developing a common understanding and a high comfort level appears to be linked to students' current understanding of the concepts or skills being evaluated. Conceptual difficulty, time allowed for practice appear to be significant factors with these goals in mind.

### Rubric Development and Practice

In adapting the self-assessment process to respective classroom contexts most teachers were concerned about time management factors. The use of the process was viewed by some as an addition to an already demanding schedule;

especially considering the impact that a two week teacher political protest and a January ice storm had on teaching schedules.

I'd never used it before, so for me it was a learning process. I found it very very time consuming and that could be because I'm new at it but I found myself worrying about the time it was taking away from other curriculum demands. (Lauren)

Time constraints were a problem for me. My students become very bored very quickly of one activity and there's no going back. Because of this we ended up sticking with our original rubric draft. The students wouldn't make any changes even when I suggested them. Part of that is my fault for telling them it was "their" rubric, not "ours." (Kate)

Time management concerns impacted the process implementation in different ways. For example, some of the teachers and students only drafted the rubrics once or twice, choosing to ignore suggestions for improvement made by the ministry consultant or by meeting participants. As a result, some of the rubrics are not as closely tied to the expectation as they could be. To demonstrate, some of the rubrics include grammar and spelling as one of the criteria, neither of which are mentioned in the expectation (see Appendix B for rubric samples).

As mentioned, time constraints also impacted practice time that students were given which in turn affected the comfort level that students felt with the process. Considerable variations occurred between classes as to the amount of

time spent using the rubric as an editing tool before student self-assessment data was collected.

We only practiced with the exemplars. I didn't have time for anything else. We went right from there to the self-assessments. (Lauren)

I observed the practice process and when I saw a difficulty, I would do a brief lesson. Then we would go back to the rubrics and practice again. (Peter)

Lack of practice might engender a reduced comfort level which may impact the reliability of the results. If students are not comfortable with the criteria and how they relate to their own writing, their judgment capacity may be negatively affected.

### Students' Attitudes Towards the Self-assessment Process and Student Ownership

Wiggins (1993), and Gillen (1991), suggest that students are more likely to participate responsibly in the experience if they view it as a valued process. If self-assessment is to be successful student must value its contribution to their learning. Again there was variation within and between classes in the degrees of ownership that students assumed over the process. According to some

teachers, students' attitude to personal achievement had an effect on the value placed on the process.

Some students just didn't understand the concepts or even seem to want to. They didn't participate well. I'd want to find a solution. (Amy)

I think it was very valuable for most of my students, especially the keen ones. Mainly because they are interested in how to improve the level of their work. My level 3's are dying to find out how to raise their work to a level 4 and the rubric provided all the necessary ticks so to speak. (Diane)

On the other hand, some of the teachers were surprised by the value that even their low performing students placed on the process. Students who hadn't been concerned about performance levels in the past were asking how they could improve certain aspects of their paragraph writing.

My guys really liked to have it all out in front of them. They'd say, "so all I have to do is this and this and I can get a level 3." I'd say, "Yes you can make your own mark right now." That really turned their crank. (Amy)

I was really impressed with the level of ownership my kids displayed. Considering that most of my students lose everything I give them, I was pleased to see that they all seemed to find a special place for their rubric and that the rubrics were being pulled out even when I hadn't asked for them. Some of my students asked me to laminate theirs! (Peter)

Student ownership may be a significant factor in the process because if students do see the process as a valuable one it is unlikely that they will assume the responsibility of self-assessment.

All of the above mentioned factors may have impacted on the validity and reliability of the self-assessment results to varying degrees. As stated earlier, there are significant differences between the classes in percentage levels of agreement thus indicating that variation within the process may impact the results. The data presented in this section underlines many the important issues relating to the technical qualities of the self-assessment process. As variations within the framework implementation occurred, so did variations arise in the technical qualities of the data. Observations about the process, the issues that emerged as well as implications for future research and practice are discussed in the following chapter.

## Chapter 5

### Discussion

This study served several purposes. First it allowed an examination of a proposed self-assessment process for formative assessment purposes. Second, it allowed the researcher to investigate the technical qualities of said process. Third it highlighted other important issues relating to the process for future self-assessment practice and investigation. Interestingly, some of these issues that arose in this study had not been mentioned by previous studies. This chapter discusses these issues in relation to process and technical issues and proposes a revised conceptual framework for future practice and investigation. It also presents the limitations of the study and suggests questions for future exploration of this problem.

The main implementation issue that the teachers identified was time management. As stated in the results section, once teachers became involved in the process, some felt that the process demanded too much time away from the regular curriculum in order to be of real use to them. As a result, some teachers omitted what may have been important steps in ensuring validity (not making any changes to the first draft of the rubric for example,) and reliability (not

providing students with enough practice time, not reviewing writing concepts.)

Interestingly, two of the teachers who were most concerned with time issues were Kate and Lauren both of whom had low inter-rater agreement. Lauren's class spent little time practicing and reviewing writing concepts on which the students were unclear. The inter-rater agreement for her class showed the lowest percentage agreement in the six class groups. On the other hand, Peter allowed for a lot of practice time, took time out to discuss writing concepts and levels of achievement with relation to the expectation and spent time with his students working through the criteria and rubric descriptions with his students. The resulting rater agreement was significantly higher than the other classes with 91% student-teacher and 64% student-peer agreement.

Considering these time concerns, one must remember that each teacher implemented the element within the same overall time frame yet the technical quality of the results varied widely. When one examines the inter-rater agreement data, it becomes evident that perceived time management issues cannot wholly account for low assessment data reliability. Although no external factors were identified by teachers as having a negative impact on the process, perhaps individual teacher's feelings of empowerment may have impacted the process. Although the researcher, repeatedly offered to act as a

resource, no teacher took advantage of her expertise outside of the monthly meetings. Unfortunately, teacher empowerment is a variable that was not examined in this instance. Taking this variable into account does not diminish the importance that the teachers placed on the issue of time. Previously, the researcher felt that time management considerations were inherent in the following elements within the conceptual framework: ensuring a common understanding and practice time. Although the element of time and how it relates to ensuring technical quality requires greater study, the researcher, in response to teacher concerns, has incorporated it into an amended conceptual framework. (See figure 4, p. 84).

Two other management issues appear to have had an impact on the self-assessment process implementation. Both the level of comfort that the teachers and the students felt with the process were issues highlighted by the teachers. All teachers felt that the process would be much more streamlined with future assessments as they and their students would have a better understanding of the process. Although, a second rubric developed by Amy's class relating to science experiments wasn't subjected to external validity and reliability checks, Amy felt that the rubric development and self-assessment progressed much more quickly as the students were now veterans in the process. Both the

students and the teacher felt better able to identify and describe important criteria relating to the designated learning expectation. All teachers felt that experience and practice would tighten the process and improve future results. These findings complement the assertions of McKinney et al. (1995), that self-assessment can help students take greater ownership for learning as it occurs.

I'm sure things would be better the second time around. It was obvious that with the little practice they had that my students self-assessments were getting better. (Ally)

It's too bad we didn't do this from the beginning of the year. Already I see improvements in my students behaviour; like now when we're writing my students look back at their rubric to check their writing. (Amy)

Besides the issues of time management and comfort level, one of the issues underlined by previous research was the importance of ensuring a common understanding of criteria and performance descriptions by the participating raters. The importance of this element cannot be overstated considering that many of the studies in the literature review highlighted this as a major consideration relating to reliability (Anderson, 1992; Mckinney et al., 1995; Higgins et al., 1994;). Though the framework attempted to account for this, most teachers felt that a common understanding was not fully achieved and is thus reflected in the accuracy of the self-assessment results. A number of

factors have been mentioned which could possibly account for this. Some of these factors include the fact that some of the teachers did not work through reliability checks with their students such as editing the rubric until language and meaning was clear to all students, some classes had a greater opportunity to practice with the rubrics than others and perhaps most significantly for the teachers was many students' lack of understanding with regard to basic paragraph construction concepts.

Working through the self-assessment process, I was surprised by how few really seemed to understand the concepts. Because I stand up at the front saying the same things over and over, I just assumed that everyone understood what I was saying. This was a good wake up call. (Dawn; Kate and Amy agreed.)

One of the more interesting results relating to reliability is the lack of agreement between student-teacher and student-peer ratings. This finding coincides with Harris et al.'s (1995) finding that students ratings of others' work were often less reliable as compared to their own self-assessment. One would assume that if students can assess their own work at a particular level of accuracy, that the same would apply to others' work. However, as is evidenced by the data, this was not the case here. When asked why this was so, the answers teachers gave were similar across the groups.

When my students did the peer evaluation, they were way off. They were rushing through it. (Lauren)

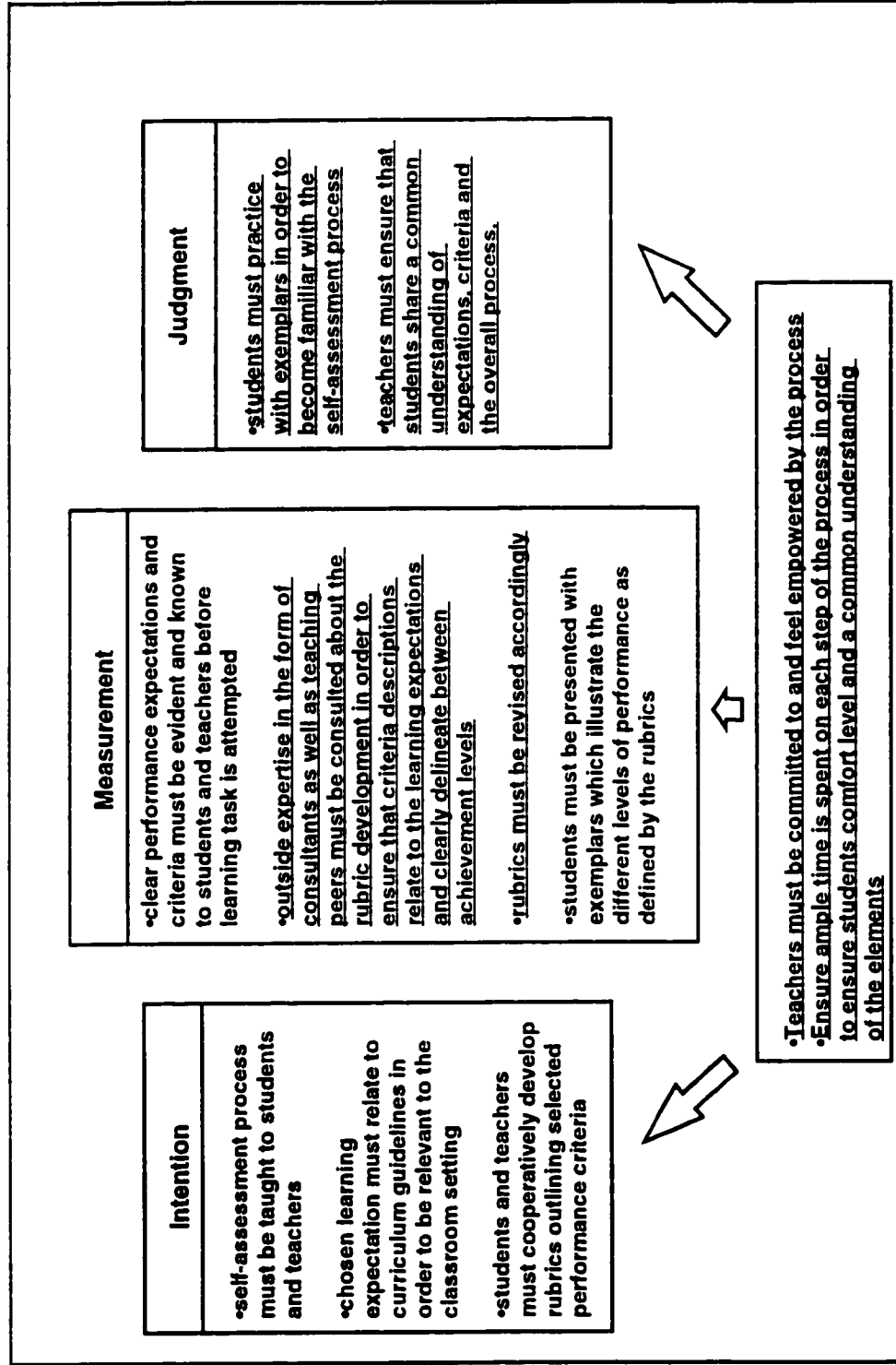
My students spent more time trying to figure out who wrote what rather than effectively assessing. If they decided that the piece that they had belonged to someone they thought was smart, they automatically gave 3's and 4's instead of looking at it. (Kate)

In order to make my students more accountable, the second time we did it I had them conference with a partner. One would have the other's and say, I don't see a topic sentence. It made them give a more honest score. They were more accountable. (Peter)

Perhaps in the first two instances commented on by teachers here, accountability was lacking thus negatively affecting the reliability of the assessment results. Peter appears to have incorporated student accountability with peer assessments with favourable results. Although the examination of the reliability of the results was done externally in order for the researcher to examine the quality of the results, it raises an interesting question regarding the importance of accountability if self-assessment is to be used in conjunction with peer assessments. However, this was not a variable investigated in this study.

One of the major findings illustrates that although user flexibility within the proposed conceptual framework may be an important aspect, implementation elements that relate to validity and reliability cannot be ignored. The examination of the classroom implementation of the proposed

self-assessment process does indicate that proposed framework elements are necessary to the process. Peter's 91% agreement rate supports the significance of the implementation elements within the conceptual framework. However, teacher level of empowerment or commitment has arisen as an important element that and is included in the revised conceptual framework. (Figure 4. p. 84.) Secondly, the issue of ensuring ample time for rubric development and student practice has also been added as an independent element.



**Figure 4:** Modified student self-assessment conceptual framework. Chronologically important elements are presented as they occurred in the assessment process. Elements that are considered ongoing and must be revisited throughout the process are underlined.

## Limitations

A number of limitations are present within this study. Ideally, the researcher would have liked to be able to observe and document an existing self-assessment process. As a definitive model for self-assessment did not exist, the researcher had to generate a process framework, thus becoming involved in the intervention. The involvement of the researcher in both in the intervention's implementation and subsequent interview may have impacted the teachers' responses. As a result of the researcher involvement, they may have been less likely to disclose negative opinions regarding the process. Secondly, the framework's flexibility allowed the teachers to implement the process as they saw fit in their classrooms. Thus there are variations in the manner that the process was completed. While some of the impacts of these variations has been presented, further study is necessary. A third limitation is the small sample size and the fact that participant selection was purposive rather than random. This presents an opportunity for bias in opinion and interpretation on the part of the teachers as they were already in favour of using alternative assessment methods and including their students in the assessment process. Fourthly, perceived external pressures were an issue. As stated previously, teachers and students had missed approximately 15 teaching

days due to a political protest and a January ice storm which resulted in school closings. Teachers felt that there was pressure to “catch up” and may not have devoted as much attention to the project as they may have in a different context. Lastly, this was a learning process for the researcher, the teachers and their students. Their lack of familiarity with the framework implementation process may have had a negative impact on the validity and the reliability of the results.

#### Implications for future research

Although this study has contributed to the existing research on self-assessment in a number of ways, many issues or questions have arisen from this study that can provide fuel for future research. First is the question, “How can this process and its technical considerations be made more classroom friendly?” Stiggins (1987) calls the teaching profession a lonely one, where teachers work in isolation rarely sharing ideas, or discussing problems with others. Considering this, external validity and reliability checks of the nature undertaken in this study are probably unfeasible for teachers without minimal external support. An investigation as to how to help teachers implement the process and produce high quality technical results without repeated intense

external intervention may perhaps encourage teachers to implement the process in their classrooms.

Related to this is the issue of teacher commitment or empowerment. Perhaps factors impacting teacher commitment to the process can be examined in order to inform and strengthen future practice.

Thirdly, a number of issues such as time management, student ownership of the process, conceptual difficulty of the expectation, and ensuring a common understanding have been highlighted as factors impacting the technical qualities of the self-assessment process. However, the degree to which each of these factors impacts the process has not been measured. Each of these issues can be examined more closely in order to determine their impact on the technical qualities of the process.

Another goal of this study was to determine whether valid, reliable self-assessments could be conducted in relation to curriculum guidelines. Does tying it to the curriculum have an impact on how the students value the process? If students were to choose their own criteria for assessment, how would this affect the accuracy of the results? Or, how can a common ground between curriculum expectations and criteria that students value relating to

their own learning be reached and how would this impact validity and reliability?

Lastly, the issue of time and its impact on the quality of rubric development, student and teacher comfort levels and the establishment of a common understanding of rubrics and the process should be examined in more detail. Unfortunately, this study did not collect pre and post practice self-assessment data thus no conclusions could be drawn as to its impact on the reliability of the results. Future investigation can isolate these concerns and examine their impact on assessment results.

### Conclusion

In conclusion, this study has only scratched the surface of self-assessment and its possibilities for use in the school classroom. The use of a conceptual framework for the self-assessment process as well as an implementation plan for this study suggests that some form of structure in the process is necessary if validity and reliability are to be ensured. The process reported here has both practical and theoretical implications. On the practical level, this study supports the introduction of self-assessment as a formative self-assessment process in the intermediate classroom. The evidence clearly states that both

teachers and students felt it was a powerful communication tool for learning and assessment. In relation to theoretical findings, although the technical quality of the results were not high, these results underlined once again the need for explicit criteria, a common understanding of the criteria and an ease of implementation. Even though further study is needed as to how to ensure that these elements are part of the self-assessment process, the findings from this study help to narrow the search. The findings in this study should help in to provide direction for future research. Lastly, from a practical perspective, the findings should help teachers plan formative self-assessments for their students while working toward valid and reliable results.

## References

Anderson, R.S. (1993), *Exploring alternative assessment: a democratic approach to student self-assessment in a reading methods class* Virginia Polytechnic Institute and State University; Blacksburg, Vi.

Archbald, D.A. & Newmann, F.F. (1988). *Beyond standardized testing*. Reston, VA: National Association of Secondary School Principals.

Betts, G.T. (1985). *Autonomous learner model for the gifted and talented*. Greely, Co: Autonomous Learning Publication and Specialists.

Black, H. & Broadfoot, P. (1982). *Keeping track of teaching*. London: Routledge and Keegan Paul Ltd.

Bloom, B.S. (1956). *Taxonomy of educational objectives*. New York, NY: David Mckay Company Inc.

Boud, D. & Falchikov, N, (1989), Quantitative studies of student self-assessment in higher education: a critical analysis of findings. *Higher Education*, 18. 529-549.

Brookfield, S. (1984). Independent adult learning. *Studies in Adult Education*, 13. (1) 15-27.

Brookfield, S. (1986). *Understanding and facilitating adult learning*. Sandfransisco: Jossey-Bass.

Butler, S. (1982), Assessing the journal: An exercise in self-assessment. *English Quarterly*, 14, (4), 75-83.

Corno, L. (1992). Encouraging student to take responsibility for learning and performance. *The Elementary School Journal*, 93, (1), 69-83.

Dumont, R.G. & Troelstrup, R.L. (1980). Exploring relationships between objective and subjective measures of intruactional outcomes. *Research in Higher Education*, 12, 1, 37-50.

Driscoll, M.P. (1994). *Psychology of learning for instruction*. Needham Heights, MA: Allyn & Bacon Inc.

Ediger, M. (1994). Measurement and evaluation. *Studies in Educational Evaluation*, 20, 169-174.

Ertmer, P.A. & Newby, T.J. (1996). The expert learner: Strategic, self-regulated, and reflective. *Instructional Science*, 24, 1-24.

Gibbons, M. & Phillips, G. (1982). Self-education: The process of life-long learning. *Canadian Journal of Education*, 7, (4), 67-86.

Gillen, M.A. (1991). The development of self-directed learning capacities: Instrumentation as one possible means. *Canadian Journal of University Continuing Education*, 17, (2), 7-20.

Gipe, J.P. (1993). Reflective thinking and growth in novices' teaching abilities. *Journal of Educational Research*, 86, (1), 52-57.

Gipps, C.V. (1995). *Beyond testing*. London: Falmer Press.

Glaser, R. (1990). Toward new models for assessment. *International Journal of Educational Research*, 14, (5), 475-83.

Goerts, J, & Betts, G.T. (1989). Center for autonomous learning. *Gifted Child Today*, 12 (5), 36-40.

Herman, J.L. & Winters, L. (1994). Portfolio research A slim collection. *Educational Leadership*, 49, (2).

Higgins, K.M., Harris, N.A. & Kuehn, L.L. (1994). Placing assessment into the hands of young children: A study of student-generated criteria and self-assessment. *Educational Assessment*, 2, (4), 309-324.

Kellaghan, T., Madaus, G.F. & Airasian, P.W. (1982). *The effects of standardized testing*. Hingham, MA: Kluwer Nijhoff Publishing:

Kincheloe, J.L. (1991). *Teachers as researchers: Qualitative inquiry as a path to empowerment*. London: Falmer Press.

Kuhns, T.M. (1994). Portfolio assessment: Making it work for the first time. *The Mathematics Teacher*, 87, (5), 332-335.

Larisey, M.M. (1994). Student self assessment A tool for learning. *Adult Learning*, 5, 9-10.

Linn, J.E. & Gronlund, M.A. (1995). *Measurement and Assessment in Teaching*. Upper Saddle, NJ: Prentice-Hall Inc.

Long, J.S. & Fransen, S.C. (1986). Evaluating the trustworthiness of self-assessments. Paper presented at The American Evaluation Association Conference. Kansas City, MO.

MacIsaac, D. & Jackson, L. (1992). Portfolio assessment: Process and outcomes. *Portfolio News*, 4, (1), 1-8

Maker, J.C. (1996). *Curriculum development and teaching strategies for gifted learners*. Austin, TX: PRO-ED Inc.

McKinney, M.O., Perkins, P.G. & Jones, W. P. (1995). Evaluating the use of self-assessment portfolios in a literacy methods class. *Reading, Research and Instruction*, 35, (1) 19-36.

Mclean, L.D. (1985). *The craft of student evaluation in Canada*. Toronto, ON: Canadian Education Association.

Meaghan, D.E. & Casas, F.R. (1995). On the testing of standards and standardized achievement testing: Panacea, placebo or pandora's box? *Interchange*, 26, (1), 33-58.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18, (2).

Ministry of Education of Ontario. (1976). *A resource guide for teachers*.

Newman, C. & Smolen, L. (1993). Portfolio assessment in our schools: Implementation, Advantages and Concerns. *Mid-Western Educational Reseracher*, 6, (1), 28-27.

Paulson, F.L & Paulson, P.R. (1992). The varieties of self-reflection. *Portfolio News*, 4, (1), 1-12.

Pike, G.R. (1995). The relationship between reports of college experience and achievement test scores. *Research in Higher Education*, 36, (1), 1-21.

Pohlmann, J.T. & Beggs, D.L. (1974). A study of the validity of self-reported measures of academic growth. *Journal of Educational Measurement*, 11, 2, 115-119.

Popham, W.J. (1995). *Classroom assessment: What teachers need to know*. Needham Heights, MA: Allyn & Bacon.

Poplin, M.S. (1988). Holistic/constructivist principles of the teaching/learning process: Implications for the field of learning disabilities. *Journal of Learning Disabilities*, 21, (7), 401-416.

Ralph, E.G. (1994). Are self-assessments accurate? Evaluating novices' teaching via triangulation. *Research in Education*, 53. 41-51.

- Rogers, C.R. (1969). *Freedom to learn*. Columbus, OH: Charles E. Merrill.
- Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-44.
- Satterly, D. (1981). *Assessment in schools*. New York, NY: Blackwell Inc.
- Spady, W.G. & Marshall, K.J. (1991). Beyond traditional outcome-based education. *Educational Leadership*, 49, (2), 67-73.
- Spandel, V. (1994). Making assessment meaningful: "Rubrics" clarify expectations, yield better feedback. *ASCD Update*, 36, (6), 4&126-127.
- Spandel, V. & Stiggins, R.J. (1990). *Creating writers: Linking assessment and writing instruction*. White Plains, NY: Longman.
- Stiggins, R.J. & Conklin, N.F. (1992). *In Teachers' hands: Investigating the practices of classroom assessment*. State University of New York Press.
- Stiggins, R.J. (1994). *Student-centered classroom assessment*. Toronto, ON: Macmillan Canada Inc.
- Treffinger, D.J. (1990). Teaching for self-directed learning: A priority for the gifted and talented. *The Gifted Child Quarterly*, 19, (1).
- Wiggins, G.P. (1993), *Assessing student performance: Exploring the purpose and limits of testing*. San Francisco, CA: Jossey Bass.

## **APPENDIX A**

### **Data collection instruments**

## Double Entry Journal Sample

### 1) Discussion of writing expectation and criteria with students

Description	Critical Reflections

**Things to think about:**

**How did you ensure that students share a common understanding of the writing expectation and the meaning of criteria?**

**What was the student response to the exercise?**

**Suggestions for improvement.**

## **Interview Guide**

**How well does this self-assessment process fit in with your own assessment philosophy?**

**How would you have assessed this expectation previously? Has the self-assessment process had any impact on the way that you assess this expectation?**

**Briefly describe the process of developing the self-assessment rubric with your class.**

**What was the level of student participation in this process?**

**How did you use exemplars in the process?**

**What kind of feedback was given to the students as they practiced using the developed rubric?**

**How did you attempt to ensure that your students shared a common understanding of the process?**

**In your own opinion, how did the students respond to this process?**

**Has the rubric helped you gain knowledge about your students? How?**

**Has the process had any impact on the way that you communicate about writing assessment with your students?**

**Did you notice any differences between students as to their level of effectiveness using the rubric?**

**Did you notice any students self-assessment patterns that emerged with the use of the rubrics?**

**If someone other than yourself was to use this rubric now with your students, do you think that the assessment results would be similar? Why?**

**What are some of the management issues related to the process?**

**Describe any strengths related to the process that you haven't mentioned.**

**Describe weaknesses.**

**Would you recommend this process to other teachers? Why?**

**Will you continue to use this process in future Language Arts assessment? Would you use it to assess other learning areas?**

**What changes, if any will you make in the process, the next time that you use it?**

## **APPENDIX B**

### **Final Version of Self-Assessment Rubrics**

**Ally: Grade 6**

**Expectation: Students will organize information to develop a central idea using well linked and developed paragraphs and a variety of sentence types**

<b>Criteria</b>	<b>Level 1</b>	<b>Level 2</b>	<b>Level 3</b>	<b>Level 4</b>
<b>Sentence Types</b>	<ul style="list-style-type: none"> <li>-only one sentence type used.</li> <li>-run-on sent.</li> <li>-several punctuation errors</li> </ul>	<ul style="list-style-type: none"> <li>-two sentence types used.</li> <li>-many minor errors in punctuation e.g. missing comma, contractions</li> </ul>	<ul style="list-style-type: none"> <li>-three sentence types used.</li> <li>-few minor punctuation errors</li> </ul>	<ul style="list-style-type: none"> <li>-many different sentence types used repeatedly.</li> <li>very few, if any punctuation errors.</li> </ul>
<b>Paragraphs</b>	<ul style="list-style-type: none"> <li>-lacking any proper order e.g. intro, body. conc.</li> <li>-no indentation</li> <li>-no central idea</li> </ul>	<ul style="list-style-type: none"> <li>-some mistakes in paragraph order</li> <li>-some indentation used</li> <li>-some paragraphs have obvious central idea</li> </ul>	<ul style="list-style-type: none"> <li>-proper paragraph order proper use of indentation</li> <li>-all paragraphs focus on a main idea</li> </ul>	<ul style="list-style-type: none"> <li>-paragraph order is logical</li> <li>-indentations and main ideas are obvious.</li> <li>-overall presentation layout of very high quality</li> </ul>

**Diane: Grade 7**

	Level 1	Level 2	Level 3	Level 4
<b>Organized information</b>	messy, no order, no titles, headings, etc.	Mostly neat, sometimes out of order, some headings missing	neat{computer}, things follow logical order headings, and titles used	extremely neat, great order, headings, titles, table of contents, footnotes used
<b>Sentence types and lengths</b>	used very few of level 3 strategies, sentences are incomplete, many run on sentences	used some of level 3 strategies few run on sentences	used all types of sentences	used all of level 3 strategies in writings of content
<b>Paragraph Structure</b>	No topic sentence, no paragraphs, sentence don't relate, no topic.	Some topic sentences, used strategies, related to sentence & paragraph structure.	Has topic sentence, all other sentences in paragraph are related to topic sentence.	Uses level 3 strategies in a variety of content
<b>Ideas &amp; Thought</b>	Strays off topic, ideas don't make sense.	Usually relates to topic, sometimes goes off target.	everything relates to main topic, well developed ideas	highly developed ideas
<b>Spelling , Voc. And Grammar</b>	many spelling errors	some spelling, grammar and punctuation errors.	Very few grammar and punctuation errors {capitals, periods, ect. }	All of level 3 and more.

## Kate: Grade 8

Criteria	Level 1	Level 2	Level 3	Level 4
Sentence Types	<ul style="list-style-type: none"> <li>-simple sentence only</li> <li>-no capitals and periods</li> </ul>	<ul style="list-style-type: none"> <li>2 types of sentences</li> <li>-very few capitals at beginning of sentences</li> </ul>	<ul style="list-style-type: none"> <li>-more than 2 sentence types used.</li> <li>-all sentences have capitals</li> <li>-proper punctuation</li> </ul>	<ul style="list-style-type: none"> <li>-more than 3 sentence types used repeatedly.</li> <li>-no punctuation errors.</li> </ul>
Organize information	<ul style="list-style-type: none"> <li>-messy</li> <li>-too many ideas put together</li> <li>-main idea missing in paragraphs</li> <li>-paragraphs in wrong order</li> </ul>	<ul style="list-style-type: none"> <li>-neater work</li> <li>-some of it hard to understand, unclear</li> </ul>	<ul style="list-style-type: none"> <li>very neat printing or writing</li> <li>-easy to follow</li> <li>-fits topic completely</li> <li>-table of contents</li> </ul>	<ul style="list-style-type: none"> <li>-computer typed</li> <li>-all of level three and glossary and footnotes</li> </ul>
Paragraph Structure	<ul style="list-style-type: none"> <li>-no paragraphs. everything is one big page</li> <li>-no indentation</li> <li>-no topic sentences</li> <li>-sentences out of order</li> </ul>	<ul style="list-style-type: none"> <li>-few indentations</li> <li>-some topic sentences</li> <li>-some paragraphs have good structure, others don't</li> </ul>	<ul style="list-style-type: none"> <li>-logical order of sentences in a paragraph</li> <li>-paragraph makes sense</li> <li>-</li> </ul>	N/C

**Grade 7/8 Self-developed Rubric**

**Expectation:** Organize information to develop a central idea, using well-linked and well-developed paragraphs as well as use a variety of sentence types and lengths.

	1	2	3	4
Indentation and main thought in first sentence	Minority of paragraphs have a topic sentence. Only a few paragraphs are clearly indented. Minority of main sentences tell the main thought for the paragraph	Most paragraphs are indented. Most paragraphs have a topic sentence. Some main sentences tell the main thought for the paragraph.	Every paragraph is indented; every paragraph has a topic sentence that has the main thought of the paragraph in it.	
Main thought backed up in body	Most paragraphs have bodies that do not back up the main thought in the first sentence.	Some paragraphs have evidence in the body that backs up the main thought in the first sentence.	All paragraphs have bodies that give evidence to back up the main point in topic sentence.	
All paragraphs talk about the main subject of the essay.	Often strays off topic. Not all paragraphs are related to main idea	Sometimes strays off topic.	All paragraphs are related to the main subject	
Use of different sentence types (e.g. declarative, interrogative etc.)	Uses only one sentence type. Improper punctuation is used.	Uses two or more different sentence types. A few sentences may have improper punctuation.	Uses three different types of sentences in piece of writing.	
Different sentence lengths.	Most sentences are similar	Paragraphs frequently have a variety of sentence types.	All paragraphs have a variety of sentence lengths.	
<del>several major</del> spelling errors	several minor spelling errors	a few minor spelling errors	practically no spelling errors	

## Amy: Grade 7/8

Criteria	Level 1	Level 2	Level 3
Proper Paragraph Form	-not indented -no elements of paragraph form	-indented -some use of proper paragraph form	-indented -proper parag. form: topic sentence, body, concluding sentence
Language Conventions	-major spelling errors -major omissions of punctuation	-several spelling errors -several punctuation and grammar errors	-few minor spelling errors -few minor errors in punctuation and grammar
Vocabulary	-small simple words	-some use of grade level vocabulary	-grade level vocab. used appropriately
Central Idea	-no clear main idea -not consistent throughout writing	-main idea expressed but not consistent throughout writing	-main idea clearly expressed -easy to identify throughout writing
Organized Info	-very disorganized -major omissions of information -difficult to follow	-some organization but information is missing or out of place	--well organized -has all the necessary info -very easy to follow

## Student-Developed Language Rubric

<i>Criteria</i>	1	2	3
<b>Proper Paragraph Form</b>	-not indented -no elements of proper paragraph form	-indented -some use of proper paragraph form	-indented -proper paragraph form: Topic sentence Body Concluding sentence
<b>Sentence Structure</b>	-simple sentences -no variety of sentence types	-mostly simple sentences -some variety of sentence types	-good variety of sentence lengths -good variety of sentence types: Interrogative Exclamatory Declarative Imperative
<b>Organized Information</b>	-writing is disorganized -major omissions of information -difficult to follow	-some organization is present -some information is missing or out of place -somewhat difficult to follow	-well organized -has all necessary information -easy to follow
<b>Central Idea</b>	-no clear main idea -not consistent throughout writing	-main idea is expressed -not consistent throughout writing	-main idea is clearly expressed -consistent throughout writing
<b>Language Conventions</b>	-major spelling errors -major omissions of punctuation	-several spelling errors occur -several errors in punctuation and grammar	-few minor spelling errors -few minor errors in punctuation and grammar
<b>Vocabulary</b>	-small, simple words	-some use of grade-level vocabulary	-grade-level vocabulary used consistently and appropriately

## **APPENDIX C**

**Synopses of Internal Process  
Within Each Classroom  
(Taken from double entry journals and interview data)**

Journal Entries	Synopsis of Student and Teacher Cooperative Self-Assessment Process	Critical Reflections
<p><u>Intention</u></p> <ul style="list-style-type: none"> <li>-Discussion of writing expectation and criteria</li> <li>-Identify criteria and create first draft of descriptive rubric</li> </ul>	<ul style="list-style-type: none"> <li>-discussed our participation in a Master's study and told students what we'd be doing. Talked about what criteria were and discussed what criteria applied to this expectation</li> <li>-students decided upon criteria to be assessed. They brainstormed as a whole class and came up with descriptions of levels 1-4 for each criteria.</li> </ul>	<ul style="list-style-type: none"> <li>-It took an extremely long time-too long.</li> <li>-The kids has difficulty focusing solely on the expectation. Wanted to include other criteria as well.</li> <li>-Some of the points the chose were ones they heard me discuss in class. I wish I hadn't told them it was their rubric because they saw it as solely theirs and didn't listen to my suggestions.</li> <li>-Small group work would have been a better way to start.</li> </ul>
<p><u>Measurement</u></p> <p>Exemplars presented and discussed. Performance description edited to align with exemplars</p> <ul style="list-style-type: none"> <li>- Practice with rubric</li> </ul>	<ul style="list-style-type: none"> <li>-Started off using rubric to judge exemplars. As they got more involved, they did not follow the rubric as well!</li> <li>-Didn't make any changes</li> <li>-We're doing a unit on food. Practiced using rubric on a creative writing piece "The Perfect Dining Experience"</li> </ul>	<p>Students wanted to include criteria such as spelling or sentence construction. I had to keep steering them back on course.</p> <ul style="list-style-type: none"> <li>- Again students got away from the rubric.</li> </ul>
<p><u>Measurement and Interpretation</u></p> <ul style="list-style-type: none"> <li>-Final changes made to rubrics if necessary</li> <li>- Use of rubric to self-assess work</li> <li>- Peer assessment of work</li> </ul>	<ul style="list-style-type: none"> <li>-We did not make any final changes due to time constraints</li> <li>- Students used rubric to assess their draft of the above assignment. They edited rough copy</li> <li>-Pieces were numbered, not named. for anonymity. Students evaluated anywhere from 2-6 pieces of work.</li> </ul>	<ul style="list-style-type: none"> <li>-The original was such a long process. The students groaned through the first one</li> <li>-For those students who are fairly bright this is a great thing to use.</li> <li>-Slower or LD students didn't use this tool as well.</li> <li>-Students worked very quickly.</li> <li>-They didn't follow the checklist as carefully as they should have.</li> <li>-Time wasted trying to figure out who wrote which story.</li> </ul>

Diane grade: 7

Process Stages	Synopsis of Student and Teacher Cooperative Self-Assessment Process	Critical Reflections
<p><b><u>Intention</u></b>            -Discussion of writing expectation and criteria            -Identify criteria and create first draft of descriptive rubric</p>	<p>-Expectation put on the board.            -General achievement levels put on board. We discussed both as a whole class.            -Talked about what criteria were.            -In cooperative groups (3 or 4 students) we designed rubrics. Combined the co-op groups' results to make a class rubric. Was very general, they hadn't identified specific criteria. Each level was like a list</p>	<p>-It took 2 forty minute language arts periods.            -Students responded well!            -They enjoyed the task.            -Students chose many of the writing strategies we had already covered regarding sentence and paragraph structure.</p>
<p><b><u>Measurement</u></b>            Exemplars presented and discussed. Performance description edited to align with exemplars            - Practice with rubric</p>	<p>-Original rubric was too general. We used the exemplars to redo it and identified specific criteria and the descriptions of each level of performance.            -Students used the rubric in various ways.            1. used rubric as a checklist for their own editing.            2. used it as a tool to help with peer editing.</p>	<p>-Students did feel ownership. They were content with their first draft and had a harder time doing the second one.            -Some students over-rate themselves. They truly believe that they are achieving level 3 work.            -Lack of concept understanding?</p>
<p><b><u>Measurement and Interpretation</u></b>            -Final changes made to rubrics if necessary            - Use of rubric to self-assess work            - Peer assessment of work</p>	<p>-Students made final changes.            -Students used rubric to assess their draft of a multi-paragraph writing assignment.</p>	<p>-They still felt comfortable with the final draft although they didn't like making changes.</p>

Ally grade: 6

Process Stages	Synopsis of Student and Teacher Cooperative Self-Assessment Process	Critical Reflections
<p><u>Intention</u></p> <ul style="list-style-type: none"> <li>-Discussion of writing expectation and criteria</li> <li>-Identify criteria and create first draft of descriptive rubric</li> </ul>	<ul style="list-style-type: none"> <li>-As a class we read over the expectation and discussed what it meant and how it could be evaluated. We used the Achievement levels standard rubric as a model.</li> <li>-Students were guided initially-used a model rubric as a guide.</li> <li>Most criteria and descriptions were are ones I would choose. Just different wording.</li> <li>Used ministry rubric to see at what level descriptions fit.</li> </ul>	<ul style="list-style-type: none"> <li>-Students generally responded well. Due to their previous experience with rubrics and expectations, students seemed comfortable with the exercise</li> <li>-Model rubrics should be discussed in one lesson and then a separate lesson for rubric development.</li> <li>-Small group or independent work should come before whole group.</li> <li>-We started with level 1...should have started with level 3</li> </ul>
<p><u>Measurement</u></p> <p>Exemplars presented and discussed. Performance description edited to align with exemplars</p> <ul style="list-style-type: none"> <li>- Practice with rubric</li> </ul>	<ul style="list-style-type: none"> <li>-Original rubric was separated into categories and rephrased for clarity.</li> <li>-Class divided into groups of 3</li> <li>Discussed which exemplars fit with levels 1-4 descriptions.</li> <li>-Main focus of discussion was level 3 quality</li> </ul>	<ul style="list-style-type: none"> <li>-Students did feel ownership to a certain extent. I did the rubric revisions and presented it to them for discussion. Students felt that their main ideas were still being used.</li> <li>-Students enjoyed the activity.</li> <li>-They wanted to get away from the rubric and include other errors in their decision.</li> <li>In the fut. I would review the skills necessary to earn a level 3.</li> </ul>
<p><u>Measurement and Interpretation</u></p> <ul style="list-style-type: none"> <li>-Final changes made to rubrics if necessary</li> <li>- Use of rubric to self-assess work</li> <li>- Peer assessment of work</li> </ul>	<ul style="list-style-type: none"> <li>-Made final draft, is quite a bit simpler.</li> <li>-When self-assessing, lots of reteaching required to ensure that all had an understanding of level 3 quality.</li> <li>-Students each reviewed a peer's work</li> </ul>	<ul style="list-style-type: none"> <li>-Final rubric easier to follow</li> <li>-Level of comfort depended on the individual. Some still struggling with what a "3" is.</li> <li>- in fut. I would let students choose the criteria that they felt comfortable evaluating to increase comfort level.</li> <li>-Peer evaluations are valuable. It helps them see the expectations from a more objective angle.</li> <li>-Most students tried to be analytical and objective.</li> <li>-I would do this again- it was a useful exercise.</li> </ul>

Process Stages	Synopsis of Student and Teacher Cooperative Self-Assessment Process	Critical Reflections
<p><b><u>Intention</u></b>                      -Discussion of writing expectation and criteria                      -Identify criteria and create first draft of descriptive rubric</p>	<ul style="list-style-type: none"> <li>-Explained to students that they would be developing a rubric for self-assess.</li> <li>-I intro'd the expectation and displayed sample rubrics. We discussed the term criteria.</li> <li>-I demo'd breaking down an expectation into criteria.</li> <li>-Students developed their own definition of criteria</li> <li>-Held a class discussion of what description of criteria should look like.</li> <li>-Broke into groups but students didn't do well</li> <li>-Finally, I used an empty rubric frame on the overhead. We discussed suggestions as the were made and chose which ones went on rubric</li> <li>-Students decided on 3 level rubric</li> </ul>	<ul style="list-style-type: none"> <li>-Students had used a self-assess. rubric previously so were aware of its function.</li> <li>-Much to my surprise, they didn't seem motivated by this opportunity, instead, they were bored.</li> <li>-I should've provided opportunities for them to practice with other expectations</li> <li>-I realized at this point how difficult it is to break expectations down into workable criteria. I should have broken the expectation down into smaller parts and developed the criteria piece by piece.</li> <li>-students disinterested. Used rewards for participation. Worked Well!</li> </ul>
<p><b><u>Measurement</u></b>                      Exemplars presented and discussed. Performance description edited to align with exemplars                      - Practice with rubric</p>	<ul style="list-style-type: none"> <li>-We used the rubric to assess the exemplars. I told them I had already assessed them and was curious to see if our assessments matched.</li> <li>-Also suggested that they should make notes as to changes that should be made in our rubric.</li> <li>-Interested students are using the rubric as a guide for their draft writing process. Some couldn't believe that was all they had to do to get a level 3?</li> <li>-I paired strong and weaker student to practice on writing samples</li> </ul>	<ul style="list-style-type: none"> <li>-I was pleased to see a sense of ownership with the rubric.</li> <li>-The students graded rigorously and nearly all had assigned the same level of assessment that I had.</li> <li>-The expectation stated plainly appears to be an incentive to conscientious students. Some expressed how they felt uncertain about what was required before.</li> <li>-Conferencing worked well as students had to assess and explain why.</li> </ul>
<p><b><u>Measurement and Interpretation</u></b>                      -Final changes made to rubrics if necessary                      - Use of rubric to self-assess work                      - Peer assessment of work</p>	<ul style="list-style-type: none"> <li>-I made the final few changes to the rubric. Showed them some of my own rubrics and explained that theirs is better.</li> <li>-Ones student commented that work would be easier if all expectations were laid out like this.</li> <li>-When self-assessing, some students rushed through it.</li> <li>-I modified assessment for IPRC'd and ESL students-they assessed 2 criteria.</li> <li>-Students conferenced when doing peer assessment, in order to help the writer strengthen piece.</li> <li>-I made them discuss why they had assessed the piece at whatever level.</li> </ul>	<ul style="list-style-type: none"> <li>-Considering that most students lose everything I assign, I was most impressed that they all found a safe spot for the rubric and pulled it out to use regularly. Some even want their rubric laminated.</li> <li>-I realized at this point that we should have done a dry run of the process.</li> <li>-Students would run through process a second time much more quickly.</li> <li>-I was pleased to see that many of my Level 1 students had improved to level 2 quality work.</li> <li>-Peer evaluations are valuable. It helps them see their work more objectively-especially for the weaker students</li> <li>-With practice, I think they would become more objective in their assessments.</li> </ul>

Amy grade: 7/8

<b>Process Stages</b>	<b>Synopsis of Student and Teacher Cooperative Self-Assessment Process</b>	<b>Critical Reflections</b>
<p><b><u>Intention</u></b>            -Discussion of writing expectation and criteria            -Identify criteria and create first draft of descriptive rubric</p>	<p>-Reviewed the evaluation process of "leveling" from 1-4            -Introduced the chosen writing expectation to the class.            -Class brainstormed the different criteria important to paragraph writing.            -In pairs, students worked at describing each level.</p>	<p>-Students generally familiar with this evaluation format.            -Students perceive levels 1-4 in terms of percentage grades (report cards)            -Students appeared familiar with most aspects of paragraph writing.            -Problem areas: differentiating between level 3 and 4.            -Most students required assistance.            -criteria "development of ideas" presented most problems.</p>
<p><b><u>Measurement</u></b>            Exemplars presented and discussed. Performance description edited to align with exemplars            - Practice with rubric</p>	<p>-Worked in pairs to use rubrics to assess exemplars. and determine examples of levels 1-4.            -Rubric became more specialized, clearer language.            -We used the rubric as a checklist for assessing and editing individual and peer work.</p>	<p>-Students enjoyed the activity.            -Help evaluating dev. of ideas was necessary.            -Difference btwn levels 3 &amp; 4 difficult for some.            -Students seem to find this activity valuable. Most students seem to be able to assess the straightforward criteria well. Not so with the more difficult ones.            -They will ask for help if unsure.</p>
<p><b><u>Measurement and Interpretation</u></b>            -Final changes made to rubrics if necessary            - Use of rubric to self-assess work            - Peer assessment of work</p>	<p>-We reviewed each of the criteria as a class            -Final draft is more specific, easier to follow.            -Students use rubric as a checklist to assess a creative writing assignment draft.            -Students each reviewed a peer's work</p>	<p>-Students seem comfortable with the first 3 criteria            -Difficulty with vocab. and dev. of ideas            -Most evaluated themselves satisfactorily.            -Some worked too quickly            -Others wanted help with the difficult criteria.            -Some took this activity very seriously, others didn't give it much thought.</p>

Lauren grade: 7

Process Stages	Synopsis of Student and Teacher Cooperative Self-Assessment Process	Critical Reflections
<p><b><u>Intention</u></b>            -Discussion of writing expectation and criteria            -Identify criteria and create first draft of descriptive rubric</p>	<p>-Introduced the chosen writing expectation to the class. Brainstormed about meaning of it.            -Defined the term criteria. Students identified criteria within expectation            -Examined sample rubric            - Reviewed criteria-whole class,            -In small groups, designed rubric for criteria            -Assimilated info from all groups to make a class rubric</p>	<p>-We needed to go over expectation many times to identify criteria.            -Some students required extra assistance            -Students enjoyed this activity. Worked over 1h. when 1/2h was planned.            -Students worked with guidance            -Without the sample rubrics, students would have had a great deal of difficulty,</p>
<p><b><u>Measurement</u></b>            Exemplars presented and discussed. Performance description edited to align with exemplars            - Practice with rubric</p>	<p>-Worked in pairs to use rubrics to assess exemplars. and determine examples of levels 1-4.            -We then worked through them as a class            -Rubric became more specialized, clearer language.            -We used the rubric as a checklist for assessing and editing different types of writing for a variety of purposes</p>	<p>-I didn't discourage any suggestions. Students decided what to change or discard in the rubric.            -It became clear while reviewing the info that a lot of the students didn't understand the writing concepts.            -It was difficult to slot this into a day because some of the students were much slower in their assessment.            -Also felt that IPRC'd and ESL students found this too difficult. Would modify somehow.</p>
<p><b><u>Measurement and Interpretation</u></b>            -Final changes made to rubrics if necessary            - Use of rubric to self-assess work            - Peer assessment of work</p>	<p>-Didn't make any changes            -Students use rubric as a checklist to assess a piece of writing from a history project that we are working on            -Students each reviewed an anonymous peer's work</p>	<p>-Students were happy with the rubric as it is            -Students have become more honest with their marks once they realized that this was not for final marking purposes.            -Because of the anonymity, I think the marks were more honest than they would have been otherwise.</p>

## **APPENDIX D**

### **Exemplars**

Grade 6: Level 1

## The American Revolution

The brittish were very mean to all the brittish people who had gone to live in the new colony because they thought that they could get a lot of money from them by making them pay lots of taxes on lots of things and this made the british people very mad, the britis solders even shot a bunch of people in boston just for throwing rocks at them and then they were taken back to england so nobody knows if they were punnished or not. only when they had the boston tea party the british stopped charging them taxes on it they didn't want the people bying it from somewher else. I think that the brittish could have been alot nicer to the colnists because they were helping to take over new contrys for england if they wer nicer the United States might not even be a seprate country today. we migt all liv together in the same place. If I had lived at that time I think I would've foght against the brittish because I wouldn't have wanted to do everything that the brittish wanted me to do and I would want to keep all my mony.

Grade 6: Level 2

### Barney the Dog

My dog Barney is very nice he has brown shaggy hair and a long waggy tail. He loves to eat potato chips and always begs when we eat them. He loves to chase after cat's and he go'es crazy if you even say "theres a cat!" Once a cat scatched him on the nose so now he does'nt get to close to them because he is afraid. He loves to sleep on my bed and sometimes he takes up so much room that I almost fell off sometimes My mom does'nt like it when he sleeps on my bed she says he leeves his hairs all over the bed which isn't helthy. When I come home from scool he goes crazy with happyness its as if he hasn't never seem me befor. me and my dad like to take him to the parc to play frisbee he loves it so much that he will play all day if he can!

We got Barney from the humane sosity. He was for month's old when we got him. He cried and cried when he had to sleep in his baskt at nite. I wood sneak downstares and spend time with him to keep him compiny. Aftr a wile he just sleeped on my bed and that is were he stays now. He is very freindly and I love him alot. Wouldn't you love to have a dog like Barny?

Grade 6: Level 3

### The Grate Escape

Reading is one of my favorite hobbies because it lets me escape to other worlds. Every world is different and the best thing is that they are all in my imagination. The thing that is great about this is that every person who reads a book can imagine the story their own way. People can look however you want and things can appear however you want them. Have you ever talked to somebody who read the same book as you but remember's different things or liked a character that you didn't like? That's okay because you are free to imagine it any way you like. This is better than tv which shows you everything and leaves nothing for you to imagine. I think kids today are losing their imaginations because of tv and computer games.

I think that schools need to teach kids that reading is fun because by grade 6 a lot of kids don't like to read books, they think it's boring or too hard. If I was trying to make kids like reading I would read books that kids like instead of the teacher always picking them. No more Anne of Green Gables or The Secret Garden. Let's read books about exciting characters like science fiction or action. I think a lot more kids would like reading if they got to read stuff like that.

Grade 6: Level 4

## Revolution

It must have been very tough to decide which side to be on. I would be mad at England because they were putting taxes on everything. The one tax I'd be most mad about would be the stamp tax imagine if everything that was printed today had a tax on it. Newspapers would go broke. On the other hand I would be afraid of losing England's protection because we wouldn't have an army big enough to fight off other countries if they invaded. I think that if I lived then I would choose to be loyal to England even though it would be very hard because all my neighbors would be against me.

The reason that I would choose to be loyal to England is that I had worked very hard to save money to go to America and wouldn't want to face tough times ahead with no support from England. Where would we get our supplies from? How would we protect us from other enemies? What if I did not like whoever was president? I would not want to leave my farm that I worked so hard to build but for my family's safety I would go north to Canada where I could live in peace and quiet.

Grade 7: Level 1

## The American Revolution

The brittish were very mean to all the brittish people who had gone to live in the new colony because they thought that they could get a lot of money from them by making them pay lots of taxes on lots of things and this made the british people very mad, the britis solders even shot a bunch of people in boston just for throwing rocks at them and then they were taken back to england so nobody knows if they were punnished or not. only when they had the boston tea party the british stopped charging them taxes on it they didn't want the people bying it from somewher else. I think that the brittish could have been alot nicer to the colnists because they were helping to take over new contrys for england if they wer nicer the United States might not even be a seprate country today. we migt all liv together in the same place. If I had lived at that time I think I would've foght against the brittish because I wouldn't have wanted to do everything that the brittish wanted me to do and I would want to keep all my mony.

Grade 7: Level 2

### The American Revolution

If I was a British Colonist living in America I would not be happy with the way that England was treating us. I would be unhappy for many reasons. Some of these reasons are: The British put too many taxes on everything, they had a stamp tax and a tea tax and tax on fabric and other stuff. Also, the colonists had no representatives in England to fight for their rights.

If I lived there I would support the idea of a revolution. I would have liked to be on that boat and dumped all that tea into the bay. That would be a really fun tea party to be at. I would also like to tar and feather the people who are loyal to England. They would deserve to get treated that way if they support such an unfair rule. I would want to pick up my gun and send the British home to England.

Grade 7: Level 3

## Revolution

It must have been very tough to decide which side to be on. I would be mad at England because they were putting taxes on everything. The one tax I'd be most mad about would be the stamp tax. Imagine if everything that was printed today had a tax on it. Newspapers would go broke. On the other hand I would be afraid of losing England's protection because we wouldn't have an army big enough to fight off other countries if they invaded. I think that if I lived then I would choose to be loyal to England even though it would be very hard because all my neighbors would be against me.

The reason that I would choose to be loyal to England is that I had worked very hard to save money to go to America and wouldn't want to face tough times ahead with no support from England. Where would we get our supplies from? How would we protect us from other enemies? What if I didn't like whoever was president? I would not want to leave my farm that I worked so hard to build but for my family's safety I would go north to Canada where I could live in peace and quiet.

Grade 7: Level 4

## The American Revolution

If I was a British colonist in the 1700's I think that I would want to remain loyal to Britain. Better the devil that you know than the one you don't. I would be unhappy about the way that England treated us but there would be a lot of problems if we broke away from them. For example: who would lead our new government? There were a lot of people at that time who had a lot of different ideas about how things should be done. Also where would we get our supplies from? Even though things were expensive and there were a lot of taxes on things at least supplies were arriving. For example: if we broke away from England they might set up ships to prevent any other countries from trading with us. I just think it would be safer to stay protected by a big country like England instead of trying to survive alone.

Being loyal would be a tough thing to do because there would be a lot of resentment from all the people who were unloyal. They did terrible things to the loyalist. Imagine being stuck in a pillory all day so that people could throw rotten fruit and vegetables at you! They even tarred and feathered them so that everybody would know that they were loyal. I would definitely think about going back to England or heading north to Canada.

Grade 8: Level 1

### Barney the Dog

My dog Barney is very nice he has brown shaggy hair and a long waggy tail. He loves to eat potato chips and always begs when we eat them. He loves to chase after cat's and he go'es crazy if you even say "theres a cat!" Once a cat scatched him on the nose so now he does'nt get to close to them because he is afraid. He loves to sleep on my bed and sometimes he takes up so much room that I almost fell off sometimes My mom does'nt like it when he sleeps on my bed she says he leeves his hairs all over the bed which isn't helthy. When I come home from scool he goes crazy with happyness its as if he hasn't never seem me befor. me and my dad like to take him to the parc to play frisbee he loves it so much that he will play all day if he can!

We got Barney from the humane sosity. He was for month's old when we got him. He cried and cried when he had to sleep in his baskt at nite. I wood sneak downstares and spend time with him to keep him compiny. Aftr a wile he just slept on my bed and that is were he stays now. He is very freindly and I love him alot. Wouldn't you love to have a dog like Barny?

Grade 8: Level 2

### My Grandfather

My grandfather has lived a very interesting life. He just celebrated his 96th birthday last month so he has seen all the cool things that have happened in the last 100 years like the invention of the car and the plane and army tanks and nuclear weapons he even fought in WW 1 he was one of those guys who would parachute behind enemy lines and try to get info about that they are doing. I wonder what I'm going to see in my lifetime.

Will we be able to live in space like on Star Trek? Are computers going to be able to think like us? Maybe I'll live on a space colony and meet all kinds of different aliens. My grandfather had his first ride in a car when he was 12 years old, a rich friend of his father's had bought one and showed it off to all his friends I think the car was a Ford. He grew up on a farm and when he was a teenager he used to take the train to Saskatchewan to work on the farms out there, I guess his dad didn't pay him so that was a better deal. All the men would sleep in a bunkhouse and there was a cook who made all their meals and if they didn't like it - tough!

My grandfather thinks that life was better 50 years ago things are changing so fast now. There were no computers 50 years ago and most people didn't go on airplanes. You would only do that if you were very rich, I don't think I'd want to live back then - no computer games, no trips to Disneyworld. It would take months to drive to Florida in one of those old Fords. He does not like computers they are too complicated for him. He says a pencil and paper and his brain are all he needs to live in this world.

I think my grandfather has lived a very interesting life because of all the things he has gotten to see. I don't think I'd want to live then because I would not want to use an outhouse in the winter. I'm very happy living in the time that we do.

Grade 8: Level 3

## Inventions

New inventions are constantly being introduced to the public. Some of the best recent inventions are the cellphone, snowboards and the dvd.

The cellphone is great because it lets you go anywhere you want and people can still get you on the phone. So you will never miss important phone calls because you're not near a phone. I think it lets people have more free time because if people can get you it doesn't matter if your at a hockygame or playing golf. You can still be available to people who need to get your opinions of things. It's also great for people who have to travel alot because they can talk to thier family's wherever they are.

Snow boards are great because they have totaly changed the world of sking. I don't know any kids who want to lern how to ski anymore-It's not Cool! There's nothing better than cruizing down a slope of deep powder coasting along on your board. My dad has even started snow boarding but he is really bad. I think he should stick to sking-he's good at that.

The dvd is a new way to watch movies. Its way better that VHS because the picture and sound is cristal clear. No more bad soundtracks or tracking problems with the picture. Soon VHS will be no longer available and we will all have to buy new VCRs but that will be great because the movies will look that much better.

People are always making new things for people to buy. These three inventions are my favourites in the last few years. I can't imagine what the future will hold for me-maybe movies will be virtual reality and we can become characters in the movie. Who knows? I can't wait to see what the futur wholds!

Grade 8: Level 4

## Mountain Biking

Mountain biking has become a very popular sport in the last few years. More mountain bikes are sold than any other kind of bicycle. One of the reasons for its popularity is that you can ride a mountain bike almost anywhere, on-road or off-road. Gravel or dirt roads are no longer off limits and special trails have even been made for mountain bikers.

Trail racing is becoming more and more popular even though it is very tough to do. Two types of races are: Descent and X-Country. In a descent race the riders start at the top of a mountain and ride down a trail. They can reach speeds of 70 km and hour while crashing down the bumpy trail! Making sharp turns and maintaining speed are the hardest things to do in a race of this type. X-country races require the riders to ride a trail that goes up and down hills. There are often big logs on the trails that they have to ride over. Sometimes the trails are too rough so the riders have to get off and run with their bikes on their shoulders.

For those of you who do not like to race mountain biking is a nice way to get out and enjoy the outdoors. You don't have to ride on busy roads because your tires allow you to ride almost anywhere. I can't think of a better way to spend a sunny summer afternoon than riding down a quiet country road and waving to the cows.