

# **THE ASSESSMENT PRACTICES OF TEACHER CANDIDATES**

by

Adam Terrance Mills

A thesis submitted to the Department of Education

in conformity with the requirements for

the degree of Master of Education

Queen's University

Kingston, Ontario, Canada

September, 2013

Copyright © Adam Terrance Mills, 2013

## Abstract

The purpose of this study is to explore what teacher candidates (n=156) believe will be their primary assessment purposes, summative assessment practices, and assessment formats in their classrooms, their values and beliefs surrounding assessment, and what contextual factors influence teacher candidates' assessment beliefs and practices. The results are placed in the context of previous research into the assessment practice of teachers and compared to the recommended assessment practices identified in research and what the Ministry of Education of Ontario expects its teachers to do through its assessment document, *Growing Success* (O.M.E., 2010). The survey instrument used in the study was composed of questions original to this research in combination with a revised version of the instrument used by McMillan (2001) and Duncan and Noonan (2007). Teacher candidates were divided into different groups based on their grade level, academic level, subject area, and B. Ed program (concurrent, or consecutive). Descriptive statistics were generated for each question by group and overall.

A Principal Components Analysis was used to reduce the 35 items in the summative assessment practices section into 5 scales for ease of interpretation. Inferential statistics (paired samples t-tests, Analysis of Variance (ANOVA) were used to determine if there were statistically significant differences between groups. The results of my research indicate that the teacher candidates report having values and beliefs supportive of the orientation towards assessments reported in the research literature and *Growing Success* (O.M.E., 2010). Teacher candidates' responses regarding purpose and format were also in alignment with practices supported in the research literature and *Growing Success* (O.M.E., 2010). In contrast, teacher candidates' uses of non-academic criteria in

making assessment decisions were not in line with recommendations found in the research literature and *Growing Success* (O.M.E., 2010). Only two statistically significant differences were noted between groups: Grades 7 and 8 teacher candidates reported using constructed response items more frequently than Grades 11 and 12 teacher candidates in their summative assessment practices; and concurrent education program teacher candidates reported their coursework as being less of an influence on their future practice than their consecutive education program peers.

## **Acknowledgements**

I should first like to credit my parents for always making educational opportunities a priority for me. The Queen's Faculty of Education Master of Education program deserves my thanks for allowing me to study there. The Assessment and Evaluation Group professors, Dr. Lyn Shulha and Dr. Liying Cheng were inspiring and helpful to my studies. To the third member of the Assessment and Evaluation Group, my supervisor Dr. Don Klinger, my thanks for your support and guidance as well as my apologies for the protracted period during which they were required. Thanks too to Mrs. Marsha Klinger for her hospitality and generous nature. Finally, my deepest appreciation is reserved for my long-suffering wife and daughter. Daphna and Rivka, thank you for everything you do to make each day a little brighter.

## Table of Contents

Abstract	ii
Acknowledgements	iv
List of Figures	vii
List of Tables	viii
List of Abbreviations	ix
Chapter 1 - Introduction	1
Purpose	4
Rationale	4
Chapter 2 - Literature Review	11
The Importance of Classroom Assessment	11
Assessment Purpose	12
Summative Assessment: Assessment <i>of</i> Learning	13
Formative Assessment: Assessment <i>for</i> and <i>as</i> Learning	14
Assessment Practice	20
Teachers and Change	21
Teachers' Current Assessment Practices	24
Factors Affecting Classroom Assessment Practice	28
Chapter 3 - Methodology	34
Overview	34
Sample	35
Survey Instrument	35
Data Collection	37
Data Analyses	38
Chapter 4 Results	42
Descriptive Statistics	42
Demographics	42
Assessment Purpose	43
Assessment Practice	48
Values and Beliefs	49
Formative and Summative Assessments Format	52

Influence of Experiences and Beliefs on Assessment	56
Factor Reduction	59
Inferential Statistics	71
Assessment Purposes	71
Assessment Practice	72
Values and Beliefs	73
Influence of Experiences and Beliefs on Assessment	74
Chapter 5 - Discussion	76
Limitations and Difficulties in Interpreting Results	76
Assessment Purpose	77
Assessment Practice	79
Values and Beliefs	82
Assessment Format	85
Influence of Experiences and Beliefs on Assessment	86
Differences between Groups	87
Final Thoughts	88
References	91
Appendix A Ethics Approval	100
Appendix B Survey Instrument	101

## **List of Figures**

Figure 1 Scree Plot for 35 Item Scale

61

## List of Tables

Table 1 Demographic Information	43
Table 2 Assessment Purpose: 5a Assess Primarily to Assign Grades?	44
Table 3 Assessment Purpose: 5b Assess Primarily to Inform Teaching?	45
Table 4 Assessment Purpose: 5c Assess Primarily to Provide Feedback?	46
Table 5 Assessment Purpose: 6a Majority of Assessments Used for Formative Purposes?	47
Table 6 Assessment Purpose: 6b Majority of Assessments Used for Summative Purposes?	48
Table 7 Descriptive Statistics: Response to Assessment Values Overall	50
Table 8 Descriptive Statistics: Response to Assessment Values by Grade	50
Table 9 Descriptive Statistics: Response to Assessment Values by Subject	51
Table 10 Descriptive Statistics: Response to Assessment Values by Level	51
Table 11 Descriptive Statistics: Response to Assessment Values by Program	52
Table 12 Descriptive Statistics: Choice of Formative Assessment Format and Rationale	53
Table 13 Descriptive Statistics: Choice of Summative Assessment Format and Rationale	54
Table 14 Influence on Assessment Practice: Beliefs Before Entering B. Ed. Program	57
Table 15 Influence on Assessment Practice: Course Work in the B. Ed. Program	58
Table 16 Influence on Assessment Practice: Teaching Practicum	59
Table 17 P.C.A. with Oblimin Rotation of Five Factor Solution for Assessment Practice Items	62
Table 18 Reliability Coefficients for Factor Structure of Assessment Practice Section	66
Table 19 Factor 1 of Assessment Practice: Motivation and Attitude	69
Table 20 Factor 2 of Assessment Practice: Higher Order Thinking	69
Table 21 Factor 3 of Assessment Practice: External Benchmarks	70
Table 22 Factor 4 of Assessment Practice: Constructed Response	70
Table 23 Factor 5 of Assessment Practice: Effort and Ability	71

## **List of Abbreviations**

ANOVA: Analysis of variance

B. Ed.: Bachelor of Education degree

P.C.A.: Principal Components Analysis

## Chapter 1

### Introduction

Debates surrounding education and, in particular, the role of teachers in helping students prepare for their futures as adult members of society have been with us for a very long time. The depiction of the trial of Socrates in Plato's *Apology* (Project Gutenberg, trans. 1891) provides historical evidence of the seriousness of disappointing a community's aspirations in regards to the education of their children. Then as now, communities tend to have very definite opinions as to the content and skills provided through education and the manner of their transmission.

In a more contemporary vein, we can add to these concerns about content and skills beliefs regarding the purpose of education in its connection to global competitiveness and the information economy (Ball, 1998; Darling-Hammond, 2005). The connections between education and the future prospects of individuals and larger communities in a competitive, global environment, and consequently the role teachers play in helping students achieve that success is coming under increasing scrutiny (Cizek, 1997; Darling-Hammond, 2005). Part of this increasing scrutiny is due to changing conceptions of effective teaching and learning and the role that assessment can play in this process. Shepard (2000) describes past conceptions of students' learning and assessment thusly, "The central ideas of social efficiency and scientific management in the curriculum circle were closely linked, respectively, to hereditarian theories of individual differences and to associationist and behaviorist learning theories" (p. 95). A student's performance on assessments was used as an indicator of aptitude and informed decisions regarding a student's suitability for further learning (Shepard, 2000).

Shepard (2000) contrasts these previous conceptions with modern approaches that focus on learning as an active process of mental construction on the part of students and a concomitant belief that all students can learn. “A commitment to equal opportunity for diverse learners means providing genuine opportunities for high quality instruction and ‘ways into’ academic curricula that are consistent with language and interaction patterns of home and community” (Shepard, 2000, p. 99). Current recommended approaches to assessment align with Shepard’s description above. Assessments that aim at uncovering students’ thinking about a subject, their level of development, and are appropriate culturally and developmentally are part and parcel of current assessment thinking as demonstrated in the research literature (Black & Wiliam, 1998a; Young & Kim, 2010) and in educational organizations’ policy documents such as the Ontario Ministry of Education’s (M. O. E.) assessment document, *Growing Success* (O.M.E., 2010).

Current assessment policy requires teachers to keep in mind the purpose to which their assessments are to be put, how the assessment purposes are reflected in the data generated by their assessment format choices, and the connections between these assessment data and teachers’ summative decisions concerning student performance. When discussing the implementation of appropriate assessment practices as indicated by research and policy it is important to keep it mind that this implementation is not a sure thing. Dye (1972) notes that the group he describes as ‘professional educators’ is comprised of several sub-groups with teachers being the numerically largest group, but politically the least significant. The struggle in effecting educational change is shared between, on the one hand, the small but influential number of educational researchers and

policy makers, and, on the other hand, the large group of classroom teachers tasked with implementing educational policies with their students.

This study aims to provide a quantitative depiction of beginning teachers' proposed future assessment practices. Such a depiction allows for a comparison between the proposed future assessment purposes and practices of beginning teachers and what research has to say about supported or recommend practices and what policy documents like *Growing Success* (O.M.E., 2010) outline as required professional practice. A focus on beginning teachers allows for assessing their 'baseline' values, beliefs and practices, before the introduction of other influencing factors such as school leadership and organization (two important factors as detailed by Young & Kim, 2010). Further, examining factors that influence teacher candidates' assessment practices can allow for an exploration of the impact these factors have on the successful implementation of sound assessment practice by beginning teachers. Finally, the influence that beginning teachers feel their previous beliefs, their Bachelor of Education (B. Ed.) coursework, and their B. Ed. practicum played in shaping their assessment beliefs and practices may provide useful information to those tasked with training teacher candidates in the future.

Given the emphasis on improved educational outcomes for students and the larger societies of which they are a part, the refinements made to our understanding of the teaching and learning process, and the role of assessment in improving learning outcomes, it behooves us to examine the extent to which the values and beliefs between the groups mentioned by Dye (1972) align. Do the practices of beginning teachers reflect the animating ideas regarding the role of assessment in teaching and learning held by educational researchers and policy makers?

## **Purpose**

The role of teachers' classroom assessment practices in improving student learning outcomes is being given increasing prominence (Cizek, 1997). However, there is still a gap amongst what researchers know about the effectiveness of different assessment purposes and practices, educational organizations' emphasis on teachers' classroom assessment policies, and the implementation of classroom assessments. Given these gaps, this study was guided by three research questions:

- What do teacher candidates imagine will be the primary assessment purposes, summative assessment practices and assessment formats used in their classroom?
- What are teacher candidates' values and beliefs surrounding assessment?
- What contextual factors influence teacher candidates' assessment beliefs and practices?

Based on these research questions, my research examined teacher candidates' reported focus on different assessment purposes and their approach to summative assessment practices. It also investigated their use of different assessment formats and their rationale for these choices. I also explored teacher candidates' values and beliefs regarding assessment and how they related to the values and beliefs underlying the research literature and the *Growing Success* document. Finally, teacher candidates' opinions regarding the relative influence of different background experiences on their assessment practices were examined.

## **Rationale**

Ball (1998) describes what he sees as the five elements of "The New Orthodoxy" shaping policy formulation in education across the globe. This new orthodoxy is

informed by concerns regarding social and economic changes. For Ball (1998), given the volatility of the economy and job prospects for the middle class, education is once again being conceived as an oligarchic good. Education now functions as a commodity to be traded for economic benefits (Ball, 1998). The colonization of education policy by the economic sphere can be seen in the New Orthodoxy; in attempts to tighten the links between education, employment, productivity and trade (the new human capital theory); in enhancing student outcomes in employment-related skills and competencies; attaining more direct control over curriculum content and assessment; in reducing costs to government in education (scaling back the state's responsibilities for public education); in increasing community input by direct involvement in school decision making and market choice (Ball, 1998).

Darling-Hammond (2005) stresses the importance of education to the future economic outcomes of today's students: "With knowledge-based work now comprising 70 percent of all new jobs, those with low levels of education can rarely find jobs at all" (p. 1). Instead of more traditional educational experiences that privileged the teacher's role in education, delivering static information to be learned by rote by students, Darling-Hammond stresses the need for a focus on new forms of teacher-student interactions and new skills. Darling-Hammond (2009) notes that the skills considered important for students to master given today's economy are not exactly new, but do have a fresh urgency. Critical thinking, problem solving, synthesis and analysis, collaboration, reflection, working independently are all skills that she feels must be included in any list of 21<sup>st</sup> Century skills. "The final dimension is the ability to learn to learn: to be able to

learn new things on one's own, to be self-guided and independent in the learning process” (Umphrey, 2009, p. 18).

The skills identified by Darling Hammond as being necessary for future learners are also promoted in *Growing Success*, the most recent assessment document published by Ontario’s Ministry of Education (O.M.E., 2010). *Growing Success* privileges the skills identified by Darling Hammond as being necessary for future learners. Aspects of Ball’s “New Orthodoxy” can be found in *Growing Success*’s explicit connections between education and future economic considerations. *Growing Success* also reflects the changing conceptions of teaching and learning identified by Shepard (2000). *Growing Success* is an example of education policy informed by economic concerns. The first three points in Ball’s (1998) ‘New Orthodoxy’ are present in the *Growing Success* document. “Education directly influences students’ life chances – and life outcomes. Today’s global, knowledge-based economy makes the ongoing work in our schools critical to our students’ success in life and to Ontario’s economic future” (O.M.E., 2010, p. 7). Employment related skills are also dealt within *Growing Success*: “the development of learning skills and work habits needed in school and in life begins early in a child’s schooling” (p. 12). The stated purpose of *Growing Success*, the provision of effective and appropriate assessment guidelines for teacher practice, gestures towards Ball’s third criterion of the New Orthodoxy, that of more control over assessment. The scope and reach of this control is incomplete, however.

My research occurred during the time when the Ontario Ministry of Education released its assessment document, *Growing Success* (O.M.E., 2010). *Growing Success* was intended to serve as a resource document to inform teachers of current conceptions

of classroom assessment, and the assessment practices that teachers should adopt in their own practice. The document has a strong focus on the use of formative assessment, reflecting the research literature that suggests formative assessment provides a means to improving student learning. *Growing Success* outlines seven fundamental principles that track closely to Shepard's (2000) depiction of the new educational assessment model. The principles dealing with assessments that are "ongoing, varied in nature, and administered over a period of time to provide multiple opportunities for students... [and] provide ongoing descriptive feedback... [and] develop student's self-assessment skills..." (p. 6) are clearly consistent with what Shepard (2000) argues current research indicates is supported practice:

From cognitive theory we have also learned that existing knowledge structures and beliefs work to enable or impede new learning, that intelligent thought involves self-monitoring and awareness about when and how to use skills, and that "expertise" develops in a field of study as a principled and coherent way of thinking and representing problems, not just as an accumulation of information. (p. 99)

Furthermore, *Growing Success* states that the use of different assessment tools is an acceptable outcome of a policy that is "committed to enabling all students to reach their potential, and to succeed. Our challenge is that every student is unique and each must have opportunities to achieve success according to his or her own interests, abilities and goals" (p. 1). The Ministry's position that all students are capable of learning and it is the obligation of teacher to facilitate this learning through the use of multiple approaches to teaching and students' demonstration of learning is commensurate with what Shepard (2000) says about the new model for education's approach to the diverse learning needs

of students in modern classrooms.

It is clear then that educational policy and educational research indicate that classroom teachers' approaches to assessment with their students be congruent with these new conceptions of teaching and learning in order to maximize students' learning outcomes (Black & William, 1998a, 1998b; Brookhart, 2007a; Cizek, 1997; Darling-Hammond, 2005; Darling-Hammond, 2009). The changing economy and social context that students will find themselves entering after schooling privileges skills different from those traditionally taught in schools (Ball, 1998; Darling-Hammond, 2005). Advances in our understanding of cognitive psychology, motivation and ability of learners all call out for teachers to embrace different approaches to classroom assessment than those they might have used in the past (Shepard, 2000). Policy documents, such as *Growing Success* (O.M.E., 2010) make it clear that teachers are expected to adopt these changes as part of their everyday classroom instruction. What is not clear is the extent to which classroom teachers (and teacher candidates who are about to start their teaching careers in their very own classrooms) are adopting these changes.

Transitions from old to newer styles of assessing are likely to be difficult. These difficulties can be exacerbated when they occur in the context of large institutions where path dependency might cause inertia in movements aimed at change. Path dependency, simply put, is the notion that past decisions affect future conditions, and once a decision is made and a system in place, it tends to perpetuate itself (Howlett, Ramesh, & Perl, 2009). Teachers and administrators whose personal classroom experiences as students were shaped by the traditional model may reject the proposed change to a new approach. Certainly Black and William (1998a, 1998b) found the uptake of newer more powerful

assessment approaches (such as formative assessment) was significantly delayed. The purpose of my research is to identify if beginning teachers have adopted the newer practices identified in the research literature and policy documents as well as the values and beliefs that inform their practice.

Blackmore (2010) employed Bourdieu's methodology to understand changes to educational policy using terms of capital and field, as well as doxa and habitus. Habitus is the common sense way of doing things possessed by individuals (and also by groups), always located in a particular field or local site/area. Each field has different knowledge or capital it privileges, providing a common sense or orthodoxy—referred to as doxa. Such a methodology is useful in identifying possible sources of conflict between the groups identified by Dye (1972) as being involved in the creation and implementation of education policies. What may seem to be 'commonsense' by teachers might not find support among researchers or policy makers. Using Bourdieu, Blackmore argues we can see how the habitus of the teacher is quickly becoming discordant with the doxa being shaped by new education policies.

Research into traditional teacher assessment practices demonstrates a habitus at odds with the modes of teaching and assessment promoted by educational assessment reformers (Black & Wiliam, 1998a, 1998b; Brookhart, 2007a; Stiggins, 2001). However, much of this same research is over a decade old. This current thesis examines whether the proposed practices of beginning classroom teachers reflect current conceptions of supported practice held by researchers and outlined in the policy document, *Growing Success* (O.M.E., 2010). This thesis investigates new B. Ed. teacher candidates' responses to questions regarding their intended future assessment purpose, assessment

format, summative assessment practice, their values and beliefs surrounding assessment, and the influence of experience on their assessment beliefs, for evidence of alignment with the current research supported practices and educational organizations policy recommendations. The role of these new teachers' values and beliefs, their educational background and their experiences in their Bachelor of Education studies are examined as potential factors affecting this alignment. The results may aid in understanding if changes in educational practice are moving to reflect a new 'commonsense.'

## Chapter 2

### Literature Review

#### The Importance of Classroom Assessment

Underpinning our understanding of the importance of classroom assessment are views (e.g., Ball, 1998; Darling-Hammond, 2005) that education must reflect what skills are required of future workers. Shepard's (2000) elucidation of the plasticity of student learning and ability dovetails nicely with a job environment that privileges adaptability over stability, and learning skills over specific content knowledge (see also Darling-Hammond, 2005). Classroom assessment practices of teachers that reflect these realities are considered powerful tools in the creation of an educated and able future work force (Darling-Hammond, 2005).

Klinger, Shulha, Mills, and Wade-Woolley (2012) credit Paul Black and Dylan Wiliam (1998a, 1998b, 1998c) for increasing the emphasis placed on teacher assessment, "Their article... proposed that properly conceived and implemented formative classroom assessment practices could directly increase student achievement" (p. 1). Black and Wiliam (1998a) stress that the new approach to assessment, "Will require significant changes in classroom practice" (p. 3). Using formative assessment to provide feedback means that, "the results have to be used to adjust teaching and learning—so a significant aspect of any programme will be the ways in which teachers do this" (Black & William, 1998a p. 3).

My research frames these changes to teachers' classroom assessment practice using three aspects: assessment purpose, assessment practice (including assessment format), and factors that affect assessment practice. The assessment purpose aspect

provides a foundation to examine refinements within the larger categories of summative and formative assessment. Research stressing the primacy of feedback is delineated. The aspect of assessment practice examines research into teachers' interactions with educational innovation. Finally, typical teacher practices according to purpose, frame of reference, assessment format and question choice, and summative assessment practice are discussed

I examine two factors that affect teachers' assessment practices: their level of expertise and their experiences. Following this, I describe the Ontario Ministry of Education's (2010) *Growing Success* document, judging its reflection of current research and supported practices. Research by McMillan (2001), and Duncan and Noonan (2007) into teachers' grading and assessment practices is presented, examining factors that affect practice. The section concludes with a discussion of key public values affecting school policy (Sergiovanni, Kelleher, McCarthy, and Wirt, 2004). Their framework provides another tool to examine the alignment between research, policy and teacher practices.

By examining these three aspects, we may determine whether misalignment between policy and beginning teacher assessment practices stem from confusion surrounding the definition and conception of assessment, misapplied or inappropriate choices for assessment practices, or background and experiential factors. The next section attempts to forestall any confusion surrounding distinctions between assessment types by examining them in detail.

### **Assessment Purpose**

At the same time as Shepard (2000) claimed changing conceptions of learning and education were taking place, distinctions between assessment purposes were being

delineated. The distinction between formative and summative was first made by Scriven (1967) in his discussion of course development. “Since Scriven drew the distinction between summative assessment and formative assessment, a gradual separation of the two into mutually exclusive entities based on the differences in functions of assessment...” (Taras, 2009, p. 59). Klinger, et al. (2012) in their review of the literature on classroom assessment trace the distinction between assessment purposes to the same time period.

The distinction between these two types of assessments is best understood as a difference in purpose, and not in the format of the assessment instrument. (Black & Wiliam, 1998b; Klinger, et al, 2011; Wiliam & Leahy, 2006; Young & Kim, 2010). This distinction is reflected in the terms used to describe these assessments: Formative assessment can be broken down into the sub-types assessment *for* learning, and assessment *as* learning (and these sub-types in turn broken down further); and summative assessment is often referred to as assessment *of* learning.

**Summative Assessment: Assessment of Learning.** Assessment *of* learning concerns judgments made of a product (test score, activity, or performance, etc.) against some set list of stated goals, criteria or standards, and other students (Taras, 2009).

Historically, assessment and evaluation were [concerned] with measuring student achievement for the purposes of grading, reporting, or promotion. Assessments were summative, and the results had relatively important consequences for students. Such assessments were largely teacher or externally driven. Performance standards were not provided and a student’s achievement was measured in

comparison to her/his peers, creating a relatively competitive norm-referenced framework within the classroom.” (Klinger, et al., 2012, p. 26-27)

Klinger, et al. (2012) contrast the historical circumstances with current summative assessment practices that ask teachers to use “a criterion-referenced framework... require[ing] stated goals, criteria or performance standards to which a student’s achievement is compared” (p. 27).

**Formative Assessment: Assessment *for* and *as* Learning.** Formative assessment also requires judgments about a product or performance to a standard or goal, but lacks the finality of summative assessments. The purpose of formative assessment is different. It seeks to provide “Information about the learning process that teachers can use for instructional decisions and students can use for improving their own performance, which motivates students” (Brookhart, 2008). Providing information about students’ progress is key to the emphasis placed on formative assessment. Black and Wiliam’s (1998a) meta-analysis identified areas that could be refined through the use of information generated by formative assessment. Klinger, et al. (2012) highlight the following areas: “curriculum design/adjustment; more impactful and targeted teacher instructional practices/content coverage; student self-monitoring and self-reflection, and peer mediated feedback” (p. 8). The information generated occurs at different times and is accessed by different users. These differences inform further divisions of formative assessment into Assessment *for* learning, and Assessment *as* learning.

Assessment *for* learning can be divided into two areas: diagnostic and formative. The distinction between them is the time at which they occur. Often occurring *prior* to a course of instruction, “diagnostic assessment principally concerns the teacher learning

about the students' learning needs" (Scaife & Wellington, 2010, p. 146). The data provided by diagnostic assessment, "allows teachers to ascertain the background knowledge a student brings with them to the start of their studies and allows for the provision of individualized and supportive instructional strategies (as well as curriculum adjustments) for students" (Klinger, et al., 2012, p. 15). The information or feedback generated by diagnostic assessments is available to and acted upon only by the teacher (Klinger, et al., 2012).

Formative assessment occurs during the course of instruction. It can be distinguished according to the level of premeditation on the part of the teacher. Formative assessment can also be distinguished by the type of information about student knowledge it provides. Yin, et al., (2008) make distinctions between formative assessments that are "on-the-fly," "planned-for-interaction," and "formal and embedded in the curriculum." "We often found that teachers proceeded through the unit only upon reaching the end to find out what and why they were teaching the activities they did... embedded assessments could be used to *signal* a unit's goal structure and give direction to teachers" (Yin, et al., 2008, p. 298). One potential hazard to the interpretation of my results is the question of whether or not teacher candidates make these same distinctions. Certainly "formal and embedded in the curriculum" assessment will be familiar to them and included in their reporting on their assessment practices; however, will teacher candidates recognize that their "on-the-fly" and "planned-for-interaction" are also powerful sources of assessment information?

Torrance and Pryor (2001) categorize teachers' formative classroom assessment practices according to whether the questions asked are 'convergent' or 'divergent.'

“Convergent assessment tasks [ask] if the learner knows/can do a specific thing... divergent assessment tasks [ask] what the student knows...” (Klinger, et al., 2012, p. 8). Convergent questions are “characterized by detailed planning” (Torrance & Pryor, 2001, p. 616). While this detailed planning is commended by Yin, et al. (2008) above, Torrance and Pryor believe convergent questioning approaches should not crowd out divergent questions:

Divergent assessment... emphasizes the learner’s understanding rather than the agenda of the assessor... the important thing is to discover what the learner knows, understands and can do... As a result, assessment is seen as accomplished jointly by the teacher and the student, and oriented more to future development rather than measurement of past or current achievement. (p. 617)

Torrance and Pryor’s (2001) research supports the research of Shepard (2000) regarding the importance of recognizing the malleability of student knowledge and learning. Darling-Hammond (2005) points out that the skills fostered by divergent questions are those identified with successful outcomes in the new global economy. The trend Shepard (2000) identifies as a focus on developing in students the skills of self-monitoring and awareness, is also identified by Klinger et al. (2012):

[as a way to] improve the quality of teacher/student interactions, and help students take active responsibility for their learning (Black and William, 1998b). This focus on student responsibility has also led to increased interest in the links between classroom assessment and metacognition, or what has since been termed “assessment as learning.” (p. 3)

This push to incorporate greater student participation in the creation and use of assessments leads to the third and final type of assessment, *Assessment as Learning*.

*Assessment as Learning* is also designed to provide information regarding student achievement on performance tasks to improve subsequent attempts but it is not teacher centered. Increasing students' responsibilities for evaluating the quality of their work provides them the opportunity to practice the skills identified by Darling-Hammond (2005) as necessary: "self-reflection", "self-regulation", "of learning how to learn." These skills fall within the realm of metacognition: "The knowledge of *and* monitoring of cognitive processes" (Salkind, 2008, p. 673). The impact of metacognition on learning and intelligence is powerful. "Metacognitive skillfulness makes a contribution to their effectiveness as a learner independently of traditionally measured intelligence... metacognitive and self-regulatory abilities are highly teachable..." (Whitebread & Pasternak, 2010, p. 673).

Self-assessment is one-way teachers can incorporate the practice of metacognitive abilities in their classrooms (McMillan, Hellsten, & Klinger, 2011). Klinger, et al. (2012) note self-assessment's "ability to motivate students to check on their own learning" (p. 21). Also important to increasing motivation is the involvement of students in the creation of the assessment (Andrade & Valtcheva, 2009; Hattie & Timperley, 2007; McMillan, et al., 2011; Sebba, et al., 2008).

Another activity with potential for increasing metacognition is peer assessment. Peer assessment allows students to provide feedback to one another. "It is an arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal status learners" (Topping, 2009, p. 21). The power of peer

feedback in improving student interest, motivation and performance is attested to by numerous studies (Gaustad, 1993; Ginsburg-Block, Rohrbeck & Fantuzzo, 2006;; Hattie, 2009; Scruggs & Mastropieri, 1998; Topping & Ehly, 1998; Topping, 2009, Tymms, Merrel, Thurston, Andor, Topping & Miller, 2011; Utley & Mortweet, 1997). However, use of these metacognitive strategies requires teachers to know and understand the content of those strategies and be comfortable using those same strategies themselves. As we will see in the second section of this literature review, this condition may not be met.

Perhaps the most important aspect of the use of formative assessment that involves students in the process is the ability to provide students with feedback about their learning. Feedback is influential in two broad areas:

through affective processes, such as increased effort, motivation, or engagement... [or] through a number of different cognitive processes, including restructuring understandings, confirming to students that they are correct or incorrect, indicating that more information is available or needed, pointing to directions student could pursue, and/or indicating alternative strategies to understand particular information. (Hattie & Timperley, 2007, p. 82)

Klinger, et al. (2012) characterize the literature on effective feedback as follows, “[it] should be focused on the task rather than the student and provide specific guidance on how to improve” (p. 9). According to Hattie and Timperley (2007), “Feedback thus is a ‘consequence’ of performance” (p. 81). The impact of this type of specific, focused feedback has been borne out by research over many years (Hattie, 1992; Hattie, 2009; Hattie & Timperly, 2007; Shepard, 2010). Brookhart (2008) provides a list of tips for

ensuring that feedback is focused. Her list underlines the switch from norm referenced assessment to criterion referenced assessment.

Multiple reviews of the literature demonstrate the importance of feedback in shaping and strengthening student motivation (Black & Wiliam, 1998a; Brookhart, 2007; Hattie & Timperley, 2007; Klinger, et al, 2012). The connection between motivation and student performance reinforces Shepard's (2000) description of the potential for students to improve and develop their skills. Dweck's (1986, 2010) findings support the idea that students' non-cognitive abilities influence their performance in school. Through an examination of Dweck's (1986, 2010) work on the epistemological beliefs of students, it becomes clear how important it is for teachers to believe in an incremental theory of intelligence that sees intellectual ability as malleable and believes that the process of seeking challenging learning goals is a pleasurable one. Even more importantly, teachers should embrace, "The position that challenge seeking and persistence are better facilitated by attempts to foster a learning goal orientation" (Dweck, 1986, p. 1045). Dweck also notes that the provision of metacognitive strategies can be powerful:

Retraining children's attributions for failure (teaching them to attribute their failures to effort or strategy instead of ability) has been shown to produce sizeable changes in persistence in the face of failure, changes that persist overtime and generalize across tasks. (Dweck, 1986, p. 1046)

Hattie and Timperley (2007) provide a framework for teachers and students to approach the creation and use of feedback in relation to metacognition:

Effective feedback must answer three major questions... Where am I going? (What are the goals?), How am I going? (What progress is being made toward the

goal?), and Where to next? (What activities need to be undertaken to make better progress?). (p. 86)

Of course it is not only students who would be wise to use these questions to guide their learning. They are important for teachers to use in monitoring their classroom assessment practices too.

### **Assessment Practice**

Current educational reforms stress the use of formative assessment practices, but are teachers' attitudes and skills amenable to this? Teachers' classroom behaviours are not necessarily focused on generating more data to provide feedback about students' understanding and performance. Concerns with adequately covering the curriculum and dealing with classroom management issues can swamp teachers' desires to assess students and then using that data for feedback. These classroom management issues are reflected in their focus when planning for classroom interaction and the way this planning for activities privileges certain approaches to assessment in their classes. "Teachers must maintain the flow of activity during a lesson or face behavioral management problems. Hence, they are faced first and foremost with deciding what activities will engage students during the lesson..." (Shavelson & Stern, 1981, p. 477).

These depictions of classroom behaviour need not be all negative. Shavelson and Stern (1981) note that teachers who are adept at classroom management use routines to effect predictability in student behaviour. "Hence a rather complex environment is rendered predictable and simplified in order for the teacher to handle its complexity" (p. 462). Classroom management techniques such as activity planning are necessary but not sufficient conditions for effective classroom assessment to take place. There would also

need to be the explicit understanding by teachers that activity selection and use in the classroom reflect the appropriate learning goals.

**Teachers and Change.** The research on the implementation of innovation and reform by teachers in their classrooms paints a less than optimistic picture. The concerns for practicality espoused by teachers and their attitudes to organizational and pedagogical change can prove impediments to educational reform (Doyle & Ponder, 1977). Desimone (2002), in her discussion of comprehensive school reforms, notes that teaching practices are resistant to change. Explanations regarding teacher resistance to innovation and reform have to account for the characteristics of teachers that might help or hinder their take up of reforms. Doyle and Ponder (1977) provide a model of teachers as “pragmatic skeptics.” This model evaluates teacher uptake of educational innovations in regards to the context teachers find themselves in, and how this shapes what teachers feel they need. Teachers use three factors to determine whether an educational reform is “practical”: “instrumentality, congruence and cost” (Doyle & Ponder, 1977, p. 7). Instrumentality refers to the innovations ability to clearly and specifically describe a procedure for use in classroom setting. The second factor, congruity, relates to the “teachers’ perceptions of their own situations” (Doyle & Ponder, 1977, p. 7). Congruity has three dimensions. The three questions teachers ask themselves when evaluating the congruity of an innovation are: is the procedure similar to the way the teacher normally teaches; who or what is the spokesperson for the innovation; is the innovation compatible with the teacher’s self-image and preferred mode of teaching (Doyle & Ponder, 1977, p. 8)?

The final factor identified by Doyle and Ponder (1977) is that of cost. “Cost may be conceptualized as a ratio between the amount of return and amount of investment. It

refers primarily to the ease with which a procedure can be implemented and the potential return for adopting an innovation” (p. 8). Another way to describe this factor is in terms of efficiency. Different instructional innovations or reforms involve different levels of investment of teachers’ time and energy. Current educational reformers focus on the role of formative feedback in providing differentiated instruction to students, a process that can at times be resource intensive. Educational resources, such as *Growing Success* (O.M.E., 2010), that promote formative assessment and individualized, personalized feedback will have to contend with the costs these policies incur for teachers. The degree to which teachers consider the costs, or the efficiency, of such interventions worthwhile will affect the policy’s adoption. Efficiency is also identified as one of four key policy goals by Sergiovanni, et al. (2004).

Brown and McIntyre (1978, 1982) discriminate between two types of educational innovation, ‘organizational’, and ‘pedagogical’ with the uptake of innovations by teachers varying between the two. Organizational innovations are adopted much more readily, than pedagogical changes. “Repeatedly, our evidence has suggested that the response of these teachers to organizational change will ultimately be acceptance of a directive from above... but response to pedagogical change will depend on other influences” (Brown & McIntyre, 1978, p. 21).

Common (1983) locates these differences in terms of power and distinguishes between two types of bureaucratic settings that influence the amount of power that individuals can exercise within those settings. The first setting (traditionally bureaucratic) is hierarchical and highly prescriptive, with enforced roles and duties (Common, 1983). In the professionally bureaucratic setting power relations are more equitable:

Relations between the teachers and administrators are ideally shaped by the notion of professional expertness and excellence... That is, the teacher is assumed to be a professional who has expertise and, because of it, is granted considerable autonomy within the organizational structure and authority to act in a self-determined fashion. (Common, 1983, p. 441)

The assumptions of expertise involved in discussions of a 'professionally bureaucratic setting' for teachers are detailed later in regards to experienced and beginning teachers. The extent to which historical educational policy trends in Ontario and the *Growing Success* (O.M.E., 2010) document reflect a professionally bureaucratic setting is discussed later, as is the importance of autonomy to teachers.

Of course, any examination of teachers' responses to educational reform must take into account the institutional aspects in which teachers are expected to work. Howlett, Ramesh, and Perl (2009) in their discussion of path dependency note that previous decisions or conditions affect future conditions. Howlett, et al. (2009) ask us to look at three different areas when attempting to explain path dependency: ideas, actors and institutions. The ideas surrounding education have changed over time. Many actors remain the same (government officials, school boards, principals and administrators, and teachers), but some new ones have been included (advocacy and lobby groups, especially those motivated by economic concerns). The institutions involved can't be seen to have changed much compared to the institutions in place before the education reformers' call to prioritize formative assessment. Overall, the degree of change among these three areas is mixed. However, treating the actors as discrete individuals (rather than as a singular role) shows much more change, particularly among teachers, as more enter and leave the

profession every day. It is these new entrants into the realm of education and their potential to effect change that has motivated my thesis research.

**Teachers' Current Assessment Practices.** Schafer and Lissitz (1987), in their paper on teachers' assessment practices, do not sugarcoat their review of literature:

We cannot help but note that this paper can be viewed as but one of a series of reviews documenting inadequacies of teacher education in the assessment area (Noll, 1955; Roeder, 1972, 1973). We hope that in fifteen years there is not yet another survey revealing little progress since the previous study, as ours has done.  
(p. 62)

Unfortunately, eleven years later, in Black and Wiliam's (1998c) review of the literature surrounding teachers' classroom assessment practices confirmed a picture of weak practice. Black & Wiliam (1998c) found that grading was overemphasized in teachers' assessments, and this same finding was repeated in Brookhart's (2007) findings. Young and Kim (2010) found in their analysis of the literature regarding teachers' use of assessments with a formative purpose that:

Teachers, however, may not view these formative uses of assessments as integral to their instruction; or, if they do, the general lack of training associated with assessments is likely to result in a struggle to do it well for all but a few individuals who might have a natural orientation towards reflection and evaluation. (p. 6)

Torrance and Pryor (2001) found similar results; assessment was typically considered a formal requirement and not an integral part of classroom practice. In situations where

formative assessment was employed, it was done inconsistently, and the use to which data generated was put was unclear.

Research into teachers' classroom assessment practices shows little movement from a norm referenced (comparing students to each other) to criterion referenced framework (Black & Wiliam, 1998c; Cizek, Fitzgerald, & Rachor, 1995/1996). This is likely a result of teachers not adequately integrating the disparate components of assessment research into a coherent assessment practice (Black & Wiliam, 1998C). Such integration is important in ensuring an effective assessment practice.

Paper and pencil quizzes and tests are still a mainstay of classroom assessment formats as are oral questions (Klinger, et al., 2012). Rubric use appears to be popular among teachers, and the benefits of such a format for the provision of feedback have been demonstrated (Andrade, Du & Wang, 2008, 2009; Andrade & Valtcheva, 2009). Unfortunately, regardless of format used, research suggests that the types of questions asked by teachers of their students are of limited quality in assessing students' understanding (Black & Wiliam, 1998a; Stiggins, 1988; Young & Kim, 2010).

Not all questions are equally valuable. The majority of questions used by teachers focus on recall and memory. Such 'low-order' questions may have a positive impact on student's basic knowledge but the use of higher order questions alone or in combination appears to be more effective to facilitate understanding and deeper learning. (Klinger, et al., 2012, pp. 11-12)

The reason for teachers' decisions to employ mainly low order questions in their classroom assessment practice may be the result of assessment literacy issues (e.g., Brookhart, 1994; Stiggins, 1988, 1991, 2001; Young & Kim, 2010). It may be more than

just teacher issues with assessment literacy. Teachers' emphasis on creating tasks to facilitate classroom management is bound to have an effect on what objectives teachers choose to cover, and how they assess students' understanding. The overreliance on lower order or convergent questions might be influenced by these classroom management concerns or by teachers' desires to generate assessments that quickly and efficiently demonstrate (if in a limited form) that their students are learning. If the latter is the case, then teachers' concerns for demonstrations of students' learning is heartening, as it aligns with research and policy supported practices. However, while the motivation for demonstrations of teacher learning are a positive, the means of achieving these demonstrations via 'lower order or convergent questions' are not in alignment with the supported practices that seek to elicit more expansive answers from students in order to generate more nuanced data on student understanding. A lack of these more nuanced data may result in less useful feedback. Furthermore, lower order questions of this variety do not target the skills identified in the literature (Darling-Hammond, 2005) as being most useful for students after they leave school and begin their careers.

Connected to assessment reference frameworks, assessment format and question type are other factors teachers take into account when conducting summative assessments:

More than half the teachers used other formal, achievement-related measures such as attendance and class participation; teachers used informal measures of achievement, such as students' answers to questions during class or their contribution to discussions; and more than half of the teachers used informal, non-achievement related measures such as student conduct and impressions of a

student's effort and teamwork in class when assigning the final grade. (Cizek, et al., 1995/1996, p. 166)

Debates frequently occur regarding the validity of including factors such as those outlined above in determining summative grades for students. While measurements of criteria that go beyond students' understandings of course content are fraught with concerns, there are also good arguments as to why these factors should be included. Dweck (1986, 2010) indicates teachers who reinforce positive behaviors increase motivation and, consequently, student performance. The perceived effort of the student is frequently cited in the research as being a factor in teachers' summative assessment decisions (Brookhart, 1994; Cizek, et al., 1995/1996; Frary, et al. 1993; Nava & Loyd, 1992; Stiggins, et al., 1989).

Perhaps the real issue is that teachers arrive at their own, individual answers to the unresolved disputes surrounding what factors should be included in generating summative grades. "In short, grades appear to consist of a potpourri of elements that vary from district to district, from teacher to teacher within a district, and even from student to student within a classroom" (Cizek, et. al, 1995/1996, p. 174). Such decisions could be influenced by teachers' values or their conceptions of the needs and abilities of the different students in their classroom. Such variability may not have been as problematic in the past, but is now given the importance placed on education as a determining factor in student's future career and life outcomes (Ball, 1998; Darling-Hammond, 2005; Shepard, 2000, 2010).

## **Factors Affecting Classroom Assessment Practice**

There is a wealth of research indicating that the experiences and expertise that teachers possess influences their assessment practices (Berliner, 2001; Darling-Hammond, 2000a, 2000b; Darling-Hammond, Chung & Frelow, 2002; Darling-Hammond, 2005; Darling-Hammond, 2009; Ferguson, 1991). Fuller (1969) was one of the first to map this area with his discussion of teachers' professional development, identifying four areas:

Pre-teaching concerns, are experienced by student teachers in their roles as pupils; survival concerns include doubts about having chosen the right occupation... teaching concerns pertain to the employment of the right teaching methods. Finally it is the concerns about pupils; achievement, motivation, and satisfaction that preoccupy the mind. (Kremer-Hayon, 1991, p. 458)

Understanding the sequencing of these concerns may shed light on the potential difficulties that beginning teachers face in their classroom assessment practices. Beginning teachers are focused on concerns emanating from their own time as students. Expectations for teachers to engage in new classroom assessment practices might be low given they are working largely from their own prior experiences. It is not until teachers have reached Fuller's (1969) third level of development that issues surrounding instruction and assessment become salient, and not until the fourth and final level of development that issues surrounding assessment are connected to student outcomes. It is in the adoption of the attitudes described by Fuller (1969) in his final level that the ideas animating the current focus on formative assessment are realized.

The ability of beginning teachers to engage in research supported assessment practices can be contrasted with the ability of experienced teachers who have a larger experiential base to draw from in their teaching assessment practices:

Teachers' anecdotes in the United States inform us that it takes three to five years until they are no longer surprised by what happens to them in their schools and classrooms... learning to teach is primarily about learning to codify knowledge so as to draw on it again (Berliner, 2004, p. 201)

Experience is a necessary but not sufficient condition for exemplary teaching (Berliner, 2001, 2004; Darling-Hammond, 2000b; Ferguson, 1991). Exemplary or expert teachers combine experience with several other habits to engender better learning outcomes for their students. Such habits include: "automaticity and routinization... sensitivity to task demands and social situations... flexibility... fast and accurate pattern-recognition capabilities" (Berlin, 2004, p. 201). This description of expertise in teaching lists skills and abilities likely to increase the ease and efficacy of formative assessment usage. Beginning teachers may lack these skills, and this spells difficulty for the operationalization of formative assessment practices in the classrooms of beginning teachers.

Another factor affecting the use of teachers' classroom assessment practices is the policy environment in which these teachers find themselves working. Anderson and Ben Jafaar (2006) in their analysis of historical policy trends in Ontario characterize current education policies as demonstrating a focus on standardized learning outcomes, teacher autonomy in their choice of instructional methods and increased accountability for uniform, standards based assessment of student performance. This description echoes a

few of the trends identified by Ball (1998) in his discussion of educational reforms, but the scope for teacher autonomy is somewhat discordant with Ball's concerns about the standardization of teacher practice. What the emphasis on teacher autonomy does reflect is Common's (1983) distinction between traditional and professionally bureaucratic settings, with the emphasis on teacher autonomy being an example of the latter.

The findings of Anderson and Ben Jafaar (2006) are of relevance to my research because of the concerns raised by Brown and McIntyre (1978, 1982) about the relative ease of uptake of organizational changes versus the lack of adoption of pedagogical changes. The current educational environment as outlined in the *Growing Success* (O.M.E., 2010) document has approached the matter of changes to assessment as a pedagogical one and not an organizational one. This is seen in the control ceded to teachers in their assessment practices in *Growing Success*. "Successful implementation [of assessment practices—especially formative assessment] will depend on the professional judgment of educators at all levels" (p. 2). As previous research recounted above has shown, a reliance on the professional judgment of teachers to implement this focus on formative assessment may be overly optimistic, particularly in the case of beginning teachers.

The educational environment represented by *Growing Success* (O.M.E., 2010) does demonstrate some positives for effective teacher practice based on educational research. *Growing Success* contains a list directing teachers to focus on creating and using assessments that demonstrate equity, transparency, support, planning, communication with students, variation, provision of feedback, and emphasis on student self-assessment. This is representative of what educational research identifies as

supported practice (Black & Wiliam, 1998a; Brookhart, 2007, Hattie & Timperley, 2007, Klinger, et al., 2011).

Moving on from discussions of the educational environment that teachers may find themselves in, research into the importance of teacher preparation finds that teacher effectiveness can be influenced by credentialing or teacher training program requirements (Darling-Hammond, 2000a, 2000b, 2009; Darling-Hammond, Chung, & Frelow, 2002).

Reviews of research over the past thirty years have concluded that even with shortcomings of current teacher education and licensing, fully prepared and certified teachers are generally better-rated and more successful with students than teachers without this preparation. (Darling-Hammond, 2000b, p. 167)

These differences in the efficacy of teacher practice based on teacher training experiences are relevant to my research. My respondents represent a sample from the larger population of teacher candidates from the Queen's University Bachelor of Education program, which has two program streams. Concurrent education candidates receive pedagogical training in conjunction with their regular course of undergraduate studies. In contrast, Consecutive education candidates only begin their formal study of education after completing at least a three-year undergraduate degree. These differences between teacher training models and the experiences they provide may mean there are significant differences in assessment practices between the two groups.

Institutional or organizational factors have been demonstrated to influence teacher assessment practice. McMillan (2001) found that grade level and academic stream did influence teacher assessment practices. It stands to reason that teachers at different grade levels and working with students of differing academic goals/abilities may employ

different assessment practices. It will be interesting to see if beginning teacher candidates' assessment practices also differ based on their reported grade level and academic stream.

McMillan (2001) and Duncan and Noonan (2007) noted differences in assessment practices amongst teachers based on subject areas. As MacMillan asks, "What accounts for these differences? It may be that... disciplines of study require different instructional techniques that promote particular grading [summative assessment] practices. *There could also be differences in the values and perspectives of teachers in each area*" [emphasis added] (p. 31). Duncan and Noonan's (2007) replication study suggested that future examinations of teachers' assessment practices should, "go beyond the influence of external factors to and investigation of the constraints teachers are under... *their internal beliefs and values, and their decision-making rationale for using assessment practices*" [emphasis added] (p. 10). Teachers' beliefs about subject matter, valid assessment techniques and their role in instruction all work together to influence their assessment practices (Young & Kim, 2010). These beliefs are important as teachers ignore ideas inimical with their own philosophies (Borko, Mayfield, Marion, Flexer, & Cumbo, 1997).

Sergiovanni, Kelleher, McCarthy, and Wirt (2004) identify the values of excellence, equity, efficiency and choice as being among the key public values affecting school policy. Excellence is defined as high standards, the ability of students to have keen understanding and the ability to perform well. Equity is defined as fair play, equal opportunity, differentiated approaches. Efficiency is defined as reduction in costs of time, money, etc. Finally, choice is defined as one's autonomy as a teacher, and the ability to personalize or individualize assessment practices. Each value can come into tension with

one of the other four values, but typically the value of excellence is juxtaposed with value of equity, and efficiency is contrasted with choice.

Through my inclusion of the values and beliefs framework created by Sergiovanni et al. (2004) I will be able to describe the values and beliefs beginning teachers feel are most important to them in relation to their assessment practices. In addition, such a framework enables an examination of how the values and beliefs of beginning teachers influence these teachers' conceptions of the purpose of assessment, their assessment practices and even the assessment tools they choose.

## Chapter 3

### Methodology

#### Overview

This research used a slightly modified version of the instrument used by both McMillan (2001) and Duncan and Noonan (2007). In both previous instances, the same three part internal structure was identified in the instrument: summative assessment practice, assessment format, and cognitive level of assessment. My research included an examination of these three parts. Questions regarding summative assessment practice asked teachers to identify the behaviours, qualities or characteristics of students they would consider in making their summative assessment decisions, what benchmarks they would use in determining students' performance on assessments, what different types of assessment formats they would use to assess, and what type of thinking skills they would be asking students to demonstrate on assessments. In addition four other areas of inquiry were added. Questions about assessment purpose, and formative and summative assessment formats were used to determine if teacher candidates' proposed assessment practices were in alignment with research recommendations and *Growing Success* (O.M.E., 2010). Questions about teacher candidates' values and beliefs were included to see if their values were in alignment with the values and beliefs underlying the research literature and the *Growing Success* document. Finally, questions about teacher background and experiences were asked of teacher candidates to determine what they felt had the most influence on their proposed future assessment practice: experiences prior to entering the program, experiences in B Ed. course work, or practicum experiences. Ethics

approval was granted by Queen's University on March 16<sup>th</sup>, 2012 (see Appendix A for letter of approval).

### **Sample**

The sample for this research was taken from the population of Intermediate/Senior teacher candidates enrolled in the Queen's University Bachelor of Education program. The size of the overall population of Intermediate and Senior B. Ed. candidates at Queen's in any academic year generally fluctuates between three and four hundred candidates. Intermediate teachers are those who teach at Grades 7 and 8, while the senior teachers work in Grades 9 to 12. The I/S program provides teacher training in all subject areas, including mathematics, science, social studies, humanities, physical education and the fine arts (visual, drama and music). There are also specialty programs in outdoor education, and technology.

Members of the population fall into one of two program streams: concurrent or consecutive. Due to the structural requirements of the B.Ed program (a minimum of three years of university education), the minimum age of members of this population is no less than twenty years of age. The majority of concurrent education candidates therefore fall in the twenty to twenty-two years of age group. The average age of students in the consecutive program is higher because many of the students (Technology, Artists in Community Education) are coming into the B.Ed program after having experienced other careers or completed full degrees.

### **Survey Instrument**

The original survey consisted of 73 items across three different sections. Of these, 23 items were deleted on the basis that they were unrelated to the Ontario educational

context. A further four items were combined into just two items for clarity. The modified form of the survey instrument proposed for this research had 61 items structured into six sections: bio-demographic information; proposed future assessment purpose; proposed summative assessment practice; assessment values and beliefs; formative and summative assessment formats; and perceived influence of teachers' experiences on their assessment practices. Respondents were asked to answer the survey as if they were teaching a course the following academic year in the same subject area, grade level, and academic stream as their most recent Ontario-based practicum experience. A copy of the survey can be found in Appendix B.

Section 1 had four items providing bio-demographic information of interest to my study: subject area, grade level, academic stream, B. Ed. program stream). The responses in this section were used to group the candidates for the purposes of subsequent statistical analyses. The second section was not in the original survey but was added in my research. This section had five questions. This section dealt with questions regarding assessment purpose. Teachers were asked what they felt the primary purpose of their assessment would be: for marks, for instructional improvement, or for giving students feedback. In addition, the last two questions in this section dealt with assessment purpose but used the technical terms for the different assessment purpose types.

Section 3 was based on the instrument used previously in the research of MacMillan (2001) and Duncan and Noonan (2007) and contained 35 items. This section asked teacher candidates to use a 6 point Likert scale to describe the frequency of which they would take different factors into consideration when generating summative grades.

The fourth section was not in the original survey but was added in my research. The 12 items for this section came from the work of Sergiovanni, et al. (2004) and Brookhart (2007). Participants' responses to these items were used to evaluate the importance of their values and beliefs on their assessment practice. The first part of this section asked students to rate each of the four values of equity, excellence, efficiency and choice (Sergiovanni, et al., 2004) on a 5-point Likert scale (1=not at all, 2=not very, 3=somewhat, 4=very, 5 extremely). The second part of this section asked respondents to rate the four values in order of importance to their proposed assessment practices.

The fifth section of the survey had two items asking beginning teachers to provide an example of an assessment format they would use for the purpose of formative assessment and the purpose of summative assessment. As described in the literature review, assessment types are distinguished by purpose, not format. This section was included to determine if teacher candidates showed preference to particular formats in either formative or summative assessments. It also allowed me to examine whether the formats chosen by teacher candidates for each assessment type really were supportive of different assessment purposes. Responses were coded and a frequency table generated. The sixth and final section, also added by me, asked three questions regarding the influence of prior experiences and beliefs, B. Ed. coursework, and the practicum experience on teachers' proposed future assessment practices.

### **Data Collection**

After receiving my ethics approval, I was permitted to begin my data collection. Vogt (2007) indicates response rates to surveys have dropped in the recent past from two-thirds to one-third (p. 91). This was born out by my experience, as the sample size for my

research was 156 (slightly less than half the total population of I/S teachers that year). Potential participants were notified of the research through an email using the Prof 155 ListServ. Posters announcing the study, providing background on the study's purpose, and detailing the manner in which volunteers could access the survey were posted on "student street" in McArthur Hall. Data collection began with the final block of classes for the year, beginning April 10, 2012 and running until April 27, 2012. Participants were provided with a website where they could access the survey. In order to further boost participation and completion rates, I provided an Information/Participation booth was set up outside of the Lecture Hall between the hours of 10:00 a.m. and 2:00 p.m. the week of April 10<sup>th</sup> -14<sup>th</sup>. Information regarding the study was made available, and paper copies of the survey were distributed. The survey took approximately 15 minutes to complete.

### **Data Analyses**

The data were first examined for outliers and/or data entry errors. Examination of the data suggested that there were only relatively few outliers, and their responses did not have a distorting effect on the data. For this reason, these responses were included in the analyses. Section one dealt with bio-demographic data and frequency counts were generated.

The data were analysed using descriptive statistics to determine the frequency, mean and standard deviation for each item in Sections 2, 3, 4, and 6 of the survey (Assessment Purpose, Assessment Practice, Values and Beliefs, and Influence of Experiences on Assessment Practice). Attention was paid to deviations from normality such as skewness or kurtosis.

The third section of my survey dealt with teacher candidates' proposed future summative assessment practices. This (excepting some minor changes) was a replication of Duncan and Noonan's 2007 study, which itself was a replication study of McMillan's (2001) research. Duncan and Noonan (2007) observed the same internal structure of the instrument in their study as that of McMillan's; however, instead of McMillan's 12-factor solution (six for grading practices, four for assessment types and two for cognitive levels of assessment), Duncan and Noonan reported a 5-factor solution. Two factors (academic enabling behaviours, and use of external benchmarks) were found for grading practices. An additional two factors were found for assessment type (constructed response assessments, and grouped quizzes/objectively scored assessments). One final factor, higher order thinking, was found for the third original construct, cognitive level of assessment.

I subjected the 35 items of the Assessment Practice scale to a Principal Components Analysis (P.C.A.) with an Oblimin rotation using SPSS 21. The suitability of the data for such a procedure was assessed. The correlation matrix contained many coefficients above .3. The Kaiser-Meyer-Olkin value was .81, above the recommended value of .6 (Pallant, 2010, p.199). The factorability of the correlation matrix was confirmed by a significant result on Bartlett's Test of Sphericity (Pallant, 2010). The initial P.C.A. identified ten components with eigenvalues greater than one. These ten components accounted for a total of 67% of the total variance. Examination of the scree plot suggested a 5-factor solution, the same as Duncan and Noonan (2007). A five factor solution accounted for 49.05% of the variance.

Given the visual evidence presented by the scree-plot and the previous findings of Duncan and Noonan (2007) I decided to continue my research using a 5-factor solution. I then ran a 5-factor P.C.A. with Oblimin rotation to reduce the total number of survey items into five factors for analyses. Reliability scores for the scales were generated using Cronbach's alpha, and items that reduced reliability were removed. Logical analysis was also used in generating the final scales. Results from the reliability tests and logical analysis resulted in a total of six items being removed across the five scales. The results of this process were then used to generate descriptive statistics on the five scale scores. These scale scores were then used as dependent variables for subsequent inferential analyses using ANOVA.

The fourth section of my survey dealt with respondents' values and beliefs surrounding their assessment practices. Their responses were analysed using paired sample t-tests to determine if there were statistically significant differences in the scores respondents' gave the different values. This was done to determine if respondents rated the importance of some value or values higher than others. In addition, the fifth section of my survey asked respondents to provide an example of assessment format for a formative assessment they would use in their future hypothetical class and an example of assessment format for a summative assessment. Responses were evaluated for cogency and coherence and coded into categories. These categories were then analysed according to frequency and reported in a frequency table. This table did not divide the reported frequencies into group or sub-group (e.g. grade or stream) as there seemed little differentiation in responses between groups.

This research examined groups differences on response scores associated with teacher candidates' answers to the bio-demographic questions in Section 1 of the survey (subject area, grade, academic level and B. Ed. program). Hence a series of 3x3x3x2 4 way ANOVAs were conducted to examine differences between the independent variables (subject area: 3; grade: 3; academic level: 3; and program stream: 2) on relevant sections of the survey: purpose, practice, values and beliefs, influence of experiences and beliefs on assessment practice. For ease of analysis (and based on Duncan and Noonan's 2007 findings) individual subjects were grouped into 1) math and science, 2) languages, humanities, and social sciences, and 3) fine arts, physical education and the technical trades. Grades (again based on Macmillan, 2001 and Duncan and Noonan, 2007) were also grouped into three units 1) 7 and 8; 2) 9 and 10; and 3) 11 and 12. Academic level also had three sub groups reflecting the divisions used in high schools in Ontario, 1) academic, 2) applied, and 3) other. The final independent variable, program stream, had two categories, concurrent and consecutive.

## Chapter 4

### Results

This chapter details the results of the analyses of the data from the survey instrument. The survey instrument was designed to provide information regarding several aspects of beginning teachers' assessment beliefs and proposed assessment practices. My research questions were:

- What do teacher candidates imagine will be the primary assessment purposes, assessment practices and assessment formats in their classroom?
- What are teacher candidates' values and beliefs surrounding assessment?
- What contextual factors influence teacher candidates' assessment beliefs and practices?

#### Descriptive Statistics

**Demographics.** The 156 responses to my survey represented just under half of the total population of Intermediate/Senior Candidates from the 2011/2012 academic year. There were a third again more consecutive candidates represented than concurrent candidates reflecting the program's greater number of consecutive candidates than concurrent ones. The survey was open to all intermediate/senior teacher candidates, but only 7 responses indicated they were answering the questions as if they were teaching intermediate classes. Almost three times as many candidates chose to respond as if they were teaching at the academic level rather than the applied or other levels. Whether this reflects the fact that more respondents to my survey had practicums at the academic level, or instead reflects respondents' preferred future classroom is unknown.

Table 1

*Demographic Information*

Factor	Label	N
Program	Concurrent	62
	Consecutive	94
Level	Academic	87
	Applied	35
	Other (open or not streamed)	34
Grades	7 and 8	7
	9 and 10	94
	11 and 12	55
Subjects	Math and Science	60
	Language/history and Social Studies	60
	Fine Arts, Phys. Ed/Technology	36
Total		156

**Assessment Purpose.** Tables 2-6 provide information regarding the respondents' proposed approach to assessment purpose in their future classrooms by group/sub-group. There were no statistically significant differences between groups as discussed later in the Inferential Statistics section. Here I examine the findings for the overall population. Questions 7a, 7b, 7c, asked students to indicate the strength to which they felt they would assess primarily to: a) assign grades, b) inform teaching, or c) provide students with feedback. These descriptors map onto assessment of learning, assessment for learning, and assessment as learning respectively. Examining the means for the responses to the first three questions, we can see that respondents considered all three assessment purposes to be somewhat to very important. This finding is confirmed by the skewness statistic, with 7a-7c demonstrating a negative skew and the kurtosis statistic indicating a

peak to the distribution of scores. While respondents indicated positive agreement with the statement “I will assess primarily to...” for all three assessment purposes, the higher scores for the questions regarding formative assessments suggest that the emphasis on formative assessment and providing students with feedback by educational researchers and educational organizations is echoed by this group of beginning teachers.

Table 2

*Assessment Purpose: 5a Assess Primarily to Assign Grades?*

Group	Sub-group	Mean (SE)	SD	Skewness (SE)	Kurtosis (SE)
Grade	7 & 8	3.71 (0.29)	0.76	0.60 (0.79)	-0.35 (1.59)
	9 & 10	3.47 (0.09)	0.88	-0.39 (0.25)	0.22 (0.49)
	11 & 12	3.42 (0.11)	0.79	-0.43 (0.32)	0.75 (0.63)
Subject	M & S	3.55 (0.10)	0.79	-0.49 (0.31)	0.93 (0.61)
	L/H/SS	3.45 (0.11)	0.87	-0.24 (0.31)	0.10 (0.61)
	A/PE/T	3.33 (0.14)	0.86	-0.44 (0.39)	0.45 (0.77)
Level	Academic	3.53 (0.09)	0.83	-0.40 (0.26)	0.17 (0.51)
	Applied	3.43 (0.14)	0.85	0.37 (0.40)	0.97 (0.78)
	Other	3.32 (0.15)	0.84	-0.37 (0.40)	0.80 (0.79)
Program	Concurrent	3.40 (0.10)	0.88	-0.60 (0.30)	0.52 (0.60)
	Consecutive	3.50 (0.08)	0.81	-0.18 (0.25)	0.18 (0.49)
Overall		3.46 (0.07)	0.84	-0.38 (0.19)	0.35 (0.39)

Table 3

*Assessment Purpose: 5b Assess Primarily to Inform Teaching?*

Group	Sub-group	Mean (SE)	SD	Skewness (SE)	Kurtosis (SE)
Grade	7 & 8	4.14 (0.34)	0.76	-0.35 (0.79)	-1.82 (1.59)
	9 & 10	3.93 (0.08)	0.88	-0.57 (0.25)	0.26 (0.49)
	11 & 12	3.89 (0.07)	0.79	-0.12 (0.32)	0.54 (0.63)
Subject	M & S	4.00 (0.82)	0.64	-0.41 (0.31)	0.92 (0.61)
	L/H/SS	3.92 (0.09)	0.70	-0.51 (0.31)	0.72 (0.61)
	A/PE/T	3.81 (0.14)	0.82	-0.27 (0.39)	-0.32 (0.77)
Level	Academic	3.99 (0.07)	0.64	-0.54 (0.26)	1.29 (0.51)
	Applied	3.80 (0.14)	0.83	-0.57 (0.40)	0.13 (0.78)
	Other	3.88 (0.13)	0.73	0.19 (0.40)	-1.03 (0.79)
Program	Concurrent	4.10 (0.10)	0.62	-0.61 (0.30)	-0.33 (0.60)
	Consecutive	3.81 (0.08)	0.74	-0.50 (0.25)	0.35 (0.49)
Overall		3.92 (0.06)	0.71	-0.45 (0.19)	0.41 (0.39)

The final two questions in this section are discussed in Tables 5 and 6, and were another way of asking teachers about their beliefs regarding the importance of assessment. Questions 7a-7c used descriptions when asking beginning teachers about their assessment purpose, but Questions 8a and 8b use more technical terms (formative vs. summative assessment) that they would have encountered during their B Ed. courses and possibly during their practicum. Again, differences between groups were not statistically significant, so I discuss the findings for the overall sample. Responses to questions 8a and 8b align with the responses given to questions 7a-7c, suggesting that respondents understand the meaning of the more technical terms, and their relation to specific assessment practices.

Table 4

*Assessment Purpose: 5c Assess Primarily to Provide Feedback?*

Group	Sub-group	Mean (SE)	SD	Skewness (SE)	Kurtosis (SE)
Grade	7 & 8	4.57 (0.20)	0.54	-0.37 (0.79)	-2.80 (1.59)
	9 & 10	4.40 (0.08)	0.72	-1.14 (0.25)	1.14 (0.49)
	11 & 12	4.49 (0.07)	0.57	-0.58 (0.32)	-0.64 (0.63)
Subject	M & S	4.33 (0.08)	0.68	-0.87 (0.31)	1.02 (0.61)
	L/H/SS	4.50 (0.08)	0.62	-1.00 (0.31)	0.00 (0.61)
	A/PE/T	4.47 (0.12)	0.70	-1.51 (0.39)	3.12 (0.77)
Level	Academic	4.46 (0.07)	0.61	-0.64 (0.26)	-0.51 (0.51)
	Applied	4.23 (0.15)	0.88	-1.03 (0.40)	0.52 (0.78)
	Other	4.62 (0.08)	0.49	-0.51 (0.40)	-1.86 (0.79)
Program	Concurrent	4.48 (0.09)	0.67	-1.28 (0.30)	1.87 (0.60)
	Consecutive	4.41 (0.07)	0.66	-0.93 (0.25)	0.76 (0.49)
Overall		4.44 (0.05)	0.67	-1.05 (0.19)	1.08 (0.39)

These results demonstrate fidelity with what the research literature and *Growing Success* (O.M.E., 2010) say about the importance of formative assessment, and particularly feedback. The alignment between responses to 5a-5c and questions 6a and 6b suggest that these teacher candidates understand the meaning of the terms formative and summative assessment, and their relation to specific assessment practices. Of course, these findings cannot guarantee that a focus on effective formative assessment will be carried out by beginning teachers in their future practice. This will be elaborated on in the discussion section.

Table 5

*Assessment Purpose: 6a Majority of Assessments Used for Formative Purposes?*

Group	Sub-group	Mean (SE)	SD	Skewness (SE)	Kurtosis (SE)
Grade	7 & 8	4.71 (0.18)	0.49	-1.23 (0.79)	-0.84 (1.59)
	9 & 10	4.05 (0.09)	0.88	-0.70 (0.25)	-0.14 (0.49)
	11 & 12	4.05 (0.09)	0.68	-0.07 (0.32)	-0.75 (0.63)
Subject	M & S	3.98 (0.12)	0.94	-0.49 (0.31)	-0.75 (0.61)
	L/H/SS	4.13 (0.09)	0.70	-0.50 (0.31)	0.28 (0.61)
	A/PE/T	4.17 (0.12)	0.74	-0.73 (0.39)	0.78 (0.77)
Level	Academic	4.00 (0.09)	0.83	-0.51 (0.26)	-0.24 (0.51)
	Applied	4.11 (0.12)	0.72	-0.17 (0.40)	-0.97 (0.78)
	Other	4.26 (0.14)	0.82	-1.23 (0.40)	1.52 (0.79)
Program	Concurrent	4.06 (0.12)	0.94	-0.75 (0.30)	-0.31 (0.60)
	Consecutive	4.10 (0.07)	0.71	-0.33 (0.25)	-0.31 (0.49)
Overall		4.08 (0.07)	0.81	-0.61 (0.19)	-0.09 (0.39)

Table 6

*Assessment Purpose: 6b Majority of Assessments Used for Summative Purposes?*

Group	Sub-group	Mean (SE)	SD	Skewness (SE)	Kurtosis (SE)
Grade	7 & 8	3.43 (0.37)	0.98	0.28 (0.79)	0.04 (1.59)
	9 & 10	3.54 (0.08)	0.74	-0.15 (0.25)	-0.23 (0.49)
	11 & 12	3.58 (0.10)	0.71	0.50 (0.32)	-0.42 (0.63)
Subject	M & S	3.55 (0.10)	0.75	0.20 (0.31)	-0.30 (0.61)
	L/H/SS	3.50 (0.10)	0.77	0.23 (0.31)	-0.29 (0.61)
	A/PE/T	3.64 (0.11)	0.68	-0.54 (0.39)	0.40 (0.77)
Level	Academic	3.56 (0.08)	0.73	0.15 (0.26)	-0.28 (0.51)
	Applied	3.49 (0.13)	0.74	-0.18 (0.40)	-0.16 (0.78)
	Other	3.59 (0.13)	0.78	0.09 (0.40)	-0.33 (0.78)
Program	Concurrent	3.56 (0.10)	0.78	0.10 (0.30)	-0.38 (0.60)
	Consecutive	3.54 (0.07)	0.71	0.03 (0.25)	-0.21 (0.49)
Overall		3.55 (0.06)	0.74	0.06 (0.19)	-0.30 (0.39)

**Assessment Practice.** The portion of the survey instrument interrogating beginning teachers' summative assessment practices is largely a replication of the one first used by McMillan (2001) and then again by Duncan and Noonan (2007) and is found in Section 3 of my survey instrument. The questions asked respondents to determine on a scale from 1 (not at all) to 6 (completely) the extent to which their summative assessments would be based on the listed criteria. The majority of the overall mean scores on the Assessment Practice question fell in Levels 3 and 4 for values of "some" and "quite a bit." The remainder of the overall mean scores fell on Level 2 or "very little". No mean scores were recorded for Level 1 ("not at all"), Level 5 ("extensively") or Level 6

(“completely”). These results suggest that respondents plan on using each of these assessment practices at least occasionally. Such findings are in agreement with research into teachers’ classroom assessment practices, but don’t necessarily reflect the recommended or supported practices in the research literature and *Growing Success* (O.M.E., 2010).

**Values and Beliefs.** Tables 7-11 provide beginning teachers’ responses regarding their values and beliefs. Using Sergiovanni et al.’s (2004) framework, respondents were asked to rate the importance of each of the four values (Equity, Efficiency, Choice, and Excellence) to their assessment practice using a 5 point Likert scale: 1= not at all, 2= not very, 3= somewhat, 4= very, 5 = extremely. All 4 values were considered at least ‘somewhat’ important by the respondents. In Table 2, the importance placed on all three values is reflected in the negative skewness statistics for each value, with scores clustering on the right side of the distribution. All four values also had a platykurtic distribution of scores, suggesting less of a peak to the scores than would be expected in a normal distribution. The highest scoring value overall was Equity ( $M=4.64$ ,  $SD=0.52$ ). The lowest scoring value overall was Efficiency, ( $M=3.06$ ,  $SD= 0.80$ ). The two values, Choice ( $M=4.19$ ,  $SD=0.63$ ) and Excellence ( $M=4.18$ ,  $SD=0.07$ ) were not that different from one another. Differences between groups were not statistically different and are discussed in the Inferential Statistics section.

Table 7

*Descriptive Statistics: Response to Assessment Values Overall*

Group	Value	Mean (SE)	SD	Skewness (SE)	Kurtosis (SE)
Overall	Equity	4.64 (0.05)	0.52	-0.97 (0.23)	-0.23 (0.46)
	Efficiency	3.06 (0.08)	0.80	0.05 (0.23)	-0.21 (0.46)
	Choice	4.19 (0.06)	0.66	-0.22 (0.23)	-0.70 (0.46)
	Excellence	4.18 (0.07)	0.7	-0.43 (0.23)	-0.22 (0.46)

Table 8

*Descriptive Statistics: Response to Assessment Values by Grade*

Group	Value	Mean (SE)	SD	Skewness (SE)	Kurtosis (SE)
Grade 7 & 8	Equity	4.70 (0.18)	0.49	-1.23 (0.80)	-0.84 (1.59)
	Efficiency	3.14 (0.26)	0.69	-0.17 (0.79)	0.34 (1.59)
	Choice	4.14 (0.14)	0.38	2.65 (0.79)	1.00 (1.59)
	Excellence	3.00 (0.11)	0.82	0.00 (0.79)	-1.20 (1.59)
Grade 9 & 10	Equity	4.63 (0.06)	0.55	-1.17 (0.25)	0.40 (0.50)
	Efficiency	3.13 (0.09)	0.91	0.18 (0.25)	-0.21 (0.50)
	Choice	4.17 (0.07)	0.67	-0.21 (0.25)	-0.76 (0.50)
	Excellence	4.12 (0.07)	0.71	-0.37 (0.25)	-0.28 (0.50)
Grade 11 & 12	Equity	4.74 (0.09)	0.63	-1.25 (0.32)	1.66 (0.63)
	Efficiency	3.09 (0.12)	0.89	-0.18 (0.32)	-0.13 (0.63)
	Choice	4.13 (0.11)	0.80	-0.47 (0.32)	-0.59 (0.63)
	Excellence	4.16 (0.09)	0.69	-0.22 (0.32)	-0.83 (0.63)

Table 9

*Descriptive Statistics: Response to Assessment Values by Subject*

Group	Value	Mean (SE)	SD	Skewness (SE)	Kurtosis (SE)
Math and Science	Equity	4.57 (0.07)	0.56	-0.86 (0.31)	-0.26 (0.61)
	Efficiency	3.20 (0.11)	0.84	-0.22 (0.31)	0.03 (0.61)
	Choice	4.03 (0.08)	0.64	-0.27 (0.31)	-0.43 (0.61)
	Excellence	4.10 (0.08)	0.63	-0.50 (0.31)	1.31 (0.61)
Languages, Humanities, Social Studies	Equity	4.67 (0.07)	0.48	-0.73 (0.31)	-1.53 (0.61)
	Efficiency	3.00 (0.12)	0.90	0.18 (0.31)	-0.10 (0.61)
	Choice	4.28 (0.09)	0.66	-0.40 (0.31)	-0.73 (0.61)
	Excellence	4.27 (0.09)	0.69	-0.42 (0.31)	-0.82 (0.61)
Arts, Phs. Ed., Tech	Equity	4.46 (0.13)	0.74	-1.45 (0.40)	2.26 (0.78)
	Efficiency	3.17 (0.16)	0.96	0.28 (0.40)	0.15 (0.78)
	Choice	4.14 (0.14)	0.85	-0.60 (0.40)	-0.50 (0.78)
	Excellence	3.94 (0.12)	0.73	0.09 (0.40)	-3.01 (0.78)

Table 10

*Descriptive Statistics: Response to Assessment Values by Level*

Group	Value	Mean (SE)	SD	Skewness (SE)	Kurtosis (SE)
Academic	Equity	4.56 (0.06)	0.59	-0.97 (0.26)	-0.03 (0.51)
	Efficiency	3.21 (0.10)	0.89	-0.12 (0.26)	0.26 (0.51)
	Choice	4.18 (0.08)	0.69	-0.26 (0.26)	-0.87 (0.51)
	Excellence	4.19 (0.08)	0.71	-0.49 (0.26)	-0.16 (0.51)
Applied	Equity	4.63 (0.08)	0.49	-0.56 (0.40)	-1.80 (0.78)
	Efficiency	3.09 (0.16)	0.92	0.30 (0.40)	0.31 (0.78)
	Choice	4.06 (0.12)	0.68	-0.07 (0.40)	-0.74 (0.78)
	Excellence	4.11 (0.11)	0.63	-0.09 (0.40)	-0.35 (0.78)
Other	Equity	4.58 (0.12)	0.66	-2.01 (0.41)	5.66 (0.80)
	Efficiency	2.91 (0.15)	0.84	0.18 (0.41)	0.21 (0.80)
	Choice	4.18 (0.13)	0.77	-0.77 (0.41)	0.57 (0.80)
	Excellence	4.00 (0.12)	0.66	0.00 (0.41)	-0.55 (0.80)

Table 11

*Descriptive Statistics: Response to Assessment Values by Program*

Group	Value	Mean (SE)	SD	Skewness (SE)	Kurtosis (SE)
Concurrent	Equity	4.60 (0.07)	0.63	-0.75 (0.30)	-0.70 (0.60)
	Efficiency	3.13 (0.10)	0.80	-0.04 (0.30)	0.07 (0.60)
	Choice	4.21 (0.09)	0.70	-0.31 (0.30)	-0.92 (0.60)
	Excellence	4.19 (0.09)	0.70	-0.58 (0.30)	0.38 (0.60)
Consecutive	Equity	4.57 (0.06)	0.62	-1.42 (0.25)	2.34 (0.50)
	Efficiency	3.11 (0.10)	0.95	0.09 (0.25)	-0.31 (0.50)
	Choice	4.12 (0.07)	0.71	-0.36 (0.25)	-0.25 (0.50)
	Excellence	4.09 (0.07)	0.67	-0.10 (0.25)	0.76 (0.50)

**Formative and Summative Assessments Format.** The fifth section of the survey asked teacher candidates to provide an example of a formative and a summative assessments that they would likely use in their future practices and to provide a rationale as to why they felt it was appropriate. Table 12 provides a frequency table for responses about formative assessment format (invalid responses not included). The frequencies for categories were not divided according to group or sub-group due to little observable difference between sub-groups and smaller and smaller frequency numbers.

126 responses were analyzed and grouped into 21 different categories. The first six most common categories will be discussed. The most common formative assessment format identified was the use of quizzes. The rationales provided by respondents reflected a purpose of providing them (the teacher) with information (assessment *for* learning) and also giving students feedback on their learning (assessment *as* learning). The emphasis on feedback echoes the findings regarding candidates' answers on the assessment purpose section and is aligned with the research literature and *Growing Success* (O.M.E., 2010).

The second most frequent response was “choice.” Responses in this group didn’t identify a specific assessment format; instead the candidates identified the need for assessments in general to reflect the opportunity for students to choose how they wished to demonstrate their learning. The rationales for this selection included concerns for equity, fairness, and the belief that by allowing students to choose the manner in which they demonstrated their learning, it would provide the teacher with a greater depth of understanding concerning student learning. This selection reinforces the responses given in the assessment values and beliefs section regarding equity. It also echoes the findings in the summative assessment practices section regarding the higher order thinking scale.

Table 12

*Descriptive Statistics: Choice of Formative Assessment Format and Rationale*

Format	Assessment Purpose and Rationale	Frequency
quiz	<i>for and as learning</i>	31
student choice	equity & know better what students know	12
Tickets out of class/exit cards	<i>as learning</i>	11
reflective/self-assessment	<i>as learning</i>	11
feedback	<i>as learning</i>	9
group tasks/jigsaw	<i>for and as learning</i>	7
checklist/rubric	<i>as learning (tell students what they need to know)</i>	5
oral feedback/questions	<i>for learning</i>	5
mock test questions	<i>for and as learning</i>	4
cooperative exercises	<i>as learning</i>	4
Group oral presentations	<i>for and as learning</i>	4
peer assessment	<i>for and as learning</i>	4
diagnostic tests	<i>for learning</i>	4
homework	<i>for and as learning</i>	3
time	provide as much time to students as they need	3
self-evaluations	<i>as learning (specifically to do with behavior)</i>	2
draft writing	<i>as learning</i>	2
teacher 1-on-1 with student	<i>for and as learning to provide differentiation</i>	2
stop/start/continue	<i>for and as learning (students' feedback to</i>	2

games	teacher)	1
summarization activities	as learning	1
	N/A	1

---

*n* = 126

Tickets out of class/exit cards and reflective/self-assessments tied for third most common responses. Tickets out of class/exit cards provide a monitoring function on student understanding in quick and simple format. Reflective/self-assessments fulfills the same purpose in a longer format. The next most common answer (much like the category ‘choice’) was feedback, an avowal of purpose, and not a specific assessment format, again showing congruity with supported and recommended assessment practices.

Table 13 provides the categories, rationales and frequency of the summative assessment practices identified by respondents. Far fewer categories were identified, as respondents demonstrated greater similarity in their answers than they had for the formative assessment format question (invalid responses not included).

Table 13

*Descriptive Statistics: Choice of Summative Assessment Format and Rationale*

Format	Assessment Purpose and Rationale	Frequency
unit test/exam	<i>of learning</i> (fair, comprehensive)	40
individual project	<i>of learning</i> (choice, creativity)	29
presentation	<i>of learning</i> (choice, creativity, sharing knowledge)	21
essay	<i>of learning</i> (choice, comprehensive)	7
group projects	<i>of learning</i> (choice)	6
choice	<i>of learning</i>	6
lab report	<i>of learning</i> (choice, in-depth)	4
rubric	<i>of learning</i> (fair, clarity)	2
time	<i>of learning</i> (equity)	2

*n* = 117

Nine different categories were identified, of which six are discussed here. Major tests, unit tests and exams were by far the most common answer. The rationales provided

by respondents were almost evenly divided between referencing that the fairness of using a common format for all students, and that this method of summative assessment was the most comprehensive in its ability to assess students' learning, particularly as it related to the use of multiple question types (multiple choice, short answer, essay, etc.). The rationale regarding the use of multiple questions aligns with issues raised by Torrance and Pryor (2001) about divergent and convergent question types and their power in improving teachers' understanding of students' learning. Multiple question types also demonstrate recommended assessment practices as discussed in *Growing Success* (O.M.E., 2010).

The second most common category was the use of individual projects. Student choice was again provided as a key rationale for this choice as was a concern for providing students an outlet for their creativity. These rationales align with responses regarding beginning teachers' beliefs about the importance of equity in their assessment practices, and also relates to the importance of motivation for students. The third most common category, presentations, shared concerns with student choice, and creativity with the category of individual projects, but added was an interest in allowing students to share their learning with the rest of the class.

The use of essays was the fourth most common answer and just like the category of major tests, the rationale given was that it provided students with choice as to what to write about, and that they were a comprehensive way to evaluate students' learning. Group projects was next, with a rationale that allowing students to work in groups gave them the choice to work on which aspect of the group project interested them the most.

This category does not reflect recommended or supported assessment practices identified in the literature or stated in *Growing Success* (O.M.E., 2010).

The final category discussed in terms of summative assessments is ‘choice’. Again we see a category that isn’t a summative assessment format but rather a purpose or concern. The concern for choice is a constant theme across assessment formats (both formative and summative) and reflects the high value that respondents place on equity in their assessment practice

**Influence of Experiences and Beliefs on Assessment.** Tables 14-16 provide the descriptive statistics for the final series of questions on the instrument, “To what extent did the following influence your beliefs about assessment: your beliefs before entering the B Ed. program; your course work in the B Ed. program; your teaching practicum.” A 5 point Likert scale was used (1=not at all, 2=not very, 3=somewhat, 4=very, 5=extremely). Results from these questions indicate that teacher candidates rated their practicum experiences as having the greatest influence on their beliefs about assessment (Overall:  $M= 4.60$ ,  $SD= 0.66$ ). Teacher candidates listed their experience of course work in the B. Ed. program as being the next most important in influencing their beliefs about assessment (Overall  $M= 3.33$ ,  $SD= 1.01$ ). This suggests that the work teacher candidates do in their course work has an influence on their assessment practices above and beyond the beliefs they brought with them into their B.Ed program (Overall  $M= 3.07$ ,  $SD= 1.00$ ). There was one statistically significant difference between groups on the question regarding the importance of B. Ed. course work. This is elaborated upon in further detail in the Inferential Statistics section.

Table 14

*Influence on Assessment Practice: Beliefs Before Entering B. Ed. Program*

Group	Sub-group	Mean (SE)	SD	Skewness (SE)	Kurtosis (SE)
Grade	7 & 8	3.43 (0.43)	1.13	0.73 (0.79)	-0.74 (1.59)
	9 & 10	3.20 (0.10)	0.96	-0.19 (0.25)	-0.19 (0.49)
	11 & 12	2.80 (0.14)	1.03	-0.22 (0.32)	-0.72 (0.63)
Subject	M & S	2.93 (0.13)	1.01	0.04 (0.31)	-0.41 (0.61)
	L/H/SS	3.12 (0.11)	0.89	-0.08 (0.31)	0.05 (0.61)
	A/PE/T	3.22 (0.19)	1.12	-0.59 (0.39)	-0.27 (0.77)
Level	Academic	3.02 (0.11)	1.01	-0.25 (0.26)	-0.46 (0.51)
	Applied	3.20 (0.16)	0.93	0.04 (0.40)	0.03 (0.78)
	Other	3.06 (0.18)	1.07	-0.12 (0.40)	-0.25 (0.78)
Program	Concurrent	3.16 (0.12)	0.94	-0.21 (0.30)	-0.01 (0.60)
	Consecutive	3.01 (0.11)	1.04	-0.14 (0.25)	-0.48 (0.49)
Overall		3.07 (0.08)	1.00	-0.18 (0.19)	-0.30 (0.39)

Table 15

*Influence on Assessment Practice: Course Work in the B. Ed. Program*

Group	Sub-group	Mean (SE)	SD	Skewness (SE)	Kurtosis (SE)
Grade	7 & 8	4.29 (0.36)	0.95	-0.76 (0.79)	-1.69 (1.59)
	9 & 10	3.29 (0.11)	1.04	-0.55 (0.25)	-0.17 (0.49)
	11 & 12	3.27 (0.12)	0.91	-0.43 (0.32)	0.03 (0.63)
Subject	M & S	3.43 (0.13)	0.98	-0.31(0.31)	-0.10 (0.61)
	L/H/SS	3.20 (0.13)	1.01	-0.63 (0.31)	0.02 (0.61)
	A/PE/T	3.36 (0.18)	1.07	-0.50 (0.39)	-0.35 (0.77)
Level	Academic	3.21 (0.11)	0.98	-0.58 (0.26)	-0.08 (0.51)
	Applied	3.54 (0.17)	0.98	-0.52 (0.40)	0.08 (0.78)
	Other	3.41 (0.19)	1.10	-0.34 (0.40)	-0.28 (0.78)
Program	Concurrent	3.03 (0.13)	1.04	-0.34 (0.30)	-0.42 (0.60)
	Consecutive	3.52 (0.10)	0.95	-0.53 (0.25)	0.19 (0.49)
Overall		3.33 (0.08)	1.01	-0.47 (0.19)	-0.13 (0.39)

Table 16

*Influence on Assessment Practice: Teaching Practicum*

Group	Sub-group	Mean (SE)	SD	Skewness (SE)	Kurtosis (SE)
Grade	7 & 8	4.70 (0.18)	0.49	-1.23 (0.79)	-0.84 (1.59)
	9 & 10	4.59 (0.07)	0.69	-2.18 (0.25)	6.85 (0.49)
	11 & 12	4.60 (0.08)	0.63	-1.80 (0.32)	4.26 (0.63)
Subject	M & S	4.55 (0.09)	0.70	-1.57(0.31)	2.21 (0.61)
	L/H/SS	4.78 (0.05)	0.42	-1.41 (0.31)	-0.01 (0.61)
	A/PE/T	4.36 (0.14)	0.83	-2.04 (0.39)	6.41 (0.77)
Level	Academic	4.66 (0.07)	0.63	-1.93 (0.26)	3.74 (0.51)
	Applied	4.46 (0.10)	0.61	-0.65 (0.40)	-0.45 (0.78)
	Other	4.59 (0.13)	0.78	-3.13 (0.40)	12.90 (0.78)
Program	Concurrent	4.58 (0.10)	0.80	-2.45 (0.30)	6.99 (0.60)
	Consecutive	4.61 (0.06)	0.55	-1.02 (0.25)	0.04 (0.49)
Overall		4.60 (0.05)	0.66	-2.06 (0.19)	6.12 (0.39)

**Factor Reduction**

The third section of the survey replicated the work of MacMillan (2001) and Duncan and Noonan (2007). McMillan (2001) and Duncan and Noonan (2007) both identified the instrument as investigating three structural components, “grading practices”, “assessment strategies”, and “cognitive levels of assessment.” McMillan (2001) used factor analysis to identify twelve components across the three original constructs of the internal structure. Six components were identified for grading practices: academic enabling behaviours, use of external benchmarks, academic achievement, use of extra credit and borderline cases, graded homework and use of zeros, and the use of

homework that is not graded. McMillan noted that the last two components were problematic given their low eigenvalues and small number of items loading on each component. Four components were identified for assessment types. These components were constructed response assessments, assessment developer, grouped quizzes and objective assessments, and major exams. The last component in this construct was also problematic.

Duncan and Noonan (2007) found a 5-factor solution. They found two factors under the structural component of grading practices: academic enabling behaviours, and the use of external benchmarks. The structural component of assessment strategies had another two factors: constructed response assessments, and grouped quizzes/objective assessments. Finally, they found one factor under the structural component of cognitive levels of assessment: higher order thinking.

I conducted a P.C.A distinguishing ten components with eigenvalues greater than 1 and rotations converged in 14 iterations. These ten components accounted for a total of 67% of the total variance. Figure 1 provides the scree plot for the factor analysis. Visual examination of Figure 1 suggested a possible 5-factor solution. These five factors accounted for 49% of the variance. A 5-factor solution coincides with the findings of Duncan and Noonan (2007). For this reason I worked with the 5-factor solution.

Figure 1

*Scree Plot for 35 Item Scale*

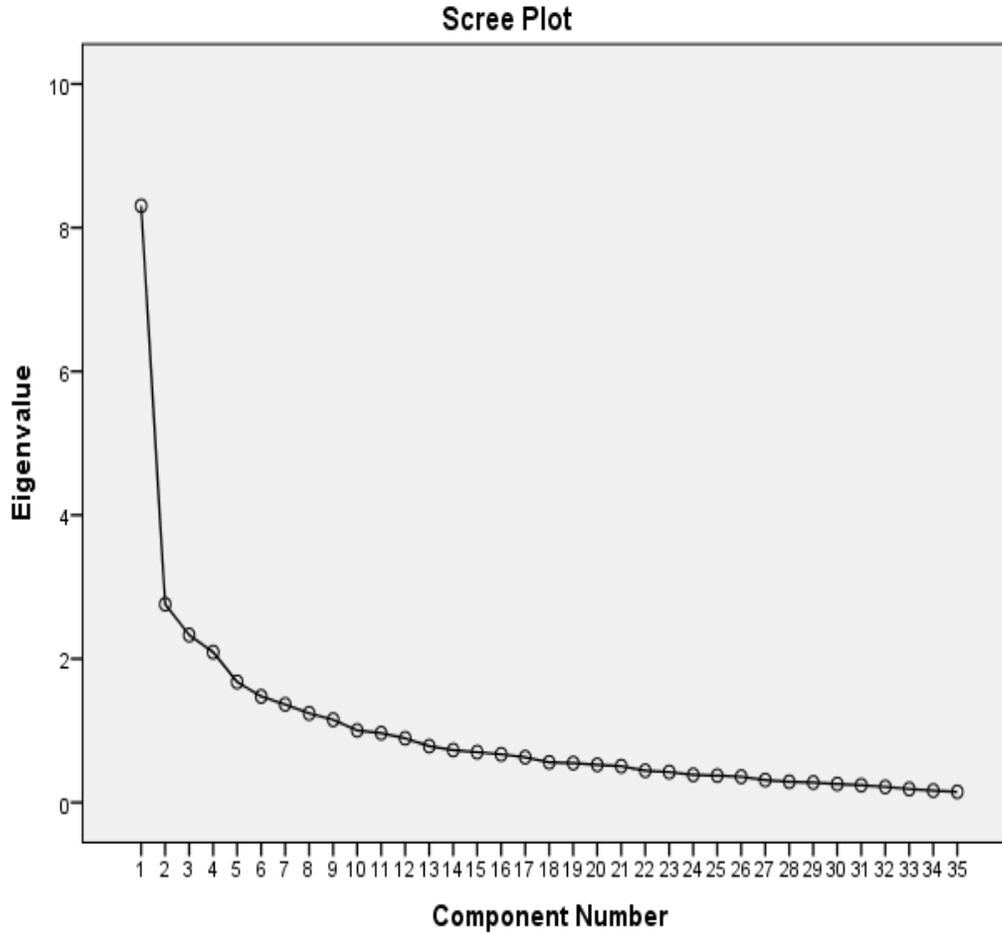


Table 17 provides the pattern matrix coefficients and communalities for the responses to my survey questions regarding beginning teacher assessment practices using a 5-factor solution with an Oblimin rotation. The results of reliability tests led to some items being removed from scales. Not all scales returned a reliability coefficient greater than .7, a hazard with scales comprised of less than ten items (Pallant, 2010).

Table 17

*P.C.A. with Oblimin Rotation of Five-Factor Solution for Assessment Practice Items*

Item	Pattern Coefficients					Communalities
	1	2	3	4	5	
Student motivation	<b>0.74</b>		0.27			0.30
How much the student tried to learn	<b>0.73</b>		0.10		0.19	0.51
Work habits and neatness	<b>0.73</b>			0.00		0.54
Student paying attention in class	<b>0.68</b>				0.14	0.49
Student seeks help	<b>0.61</b>		0.27		0.10	0.61
Class participation	<b>0.58</b>			-0.31		0.53
Effort, improvement, behavior in borderline cases	<b>0.55</b>	0.23				0.40
Only academic performance	<b>0.53</b>	0.38	0.26		0.27	0.36
Student attitudes	<b>0.47</b>	-0.11	0.13		0.40	0.59
Homework completed	<b>0.47</b>	0.15	0.21		0.11	0.54
Care in completing assignments	<b>0.46</b>	0.38	0.10		0.11	0.65
Individual projects		<b>0.76</b>				0.41
Assessments of reasoning or application		<b>0.71</b>	-0.13			0.42
Mastery of specific objectives		<b>0.60</b>				0.53
Authentic assessments (relevant outside school)	0.27	<b>0.57</b>	-0.19			0.53
Assessments of recall or understanding		<b>0.43</b>	0.25		0.13	0.33
Improvement by initially low performing student	0.27	<b>0.31</b>	-0.10	-0.19	0.14	0.54
Compared to previous students		-0.22	<b>0.76</b>			0.61
Compared to students that semester			<b>0.74</b>			0.40
Set scale of percentage correct	0.18		<b>0.59</b>	0.18		0.39
Performance on quizzes	0.10	0.10	<b>0.42</b>		0.19	0.57
How the student did on homework	0.36	0.16	<b>0.37</b>	-0.12		0.51
Major exams or compositions	-0.29	0.22	<b>0.35</b>	-0.15	0.25	0.53
Oral presentations			-0.14	<b>-0.77</b>		0.70
Essay type tests	-0.24		0.16	<b>-0.72</b>		0.33
Student responses to oral questions during class	0.32			<b>-0.68</b>		0.53
Performance assessments		0.14	-0.11	<b>-0.68</b>		0.43
Projects completed in teams	0.16	0.19		<b>-0.52</b>		0.53
Objective assessments	-0.17		0.43	<b>-0.43</b>	-0.17	0.67
High student effort	0.34	-0.18			<b>0.65</b>	0.54
Improvement of performance		0.10	-0.20	-0.21	<b>0.63</b>	0.41
Degree of effort of low ability student	0.25	0.11			<b>0.62</b>	0.57
Overall ability level of the student		0.18			<b>0.55</b>	0.31
Including zeros	-0.23			0.16	<b>0.52</b>	0.44
Low student effort	0.17	-0.27		-0.33	<b>0.52</b>	0.43

Note: major loading for each item

bolded

a. Rotation converged in 14 iterations.

Factor 1 dealt with how students' motivation and attitude influence beginning teachers' future grading practices. This was consistent with Duncan and Noonan's (2007) results finding academic enabling behaviors as one of two factors under the structural component of grading practices. In my results, one question "using only academic performance" that loaded on to Factor 1 in the P.C.A. had to be removed due to its negative influence on the inter-item correlation matrix. Removing this item left the ten-item scale with a reliability coefficient of 0.89, a robust result. This seems logical as "using only academic performance" bears little to no conceptual relationship to the other items in this scale, all of which deal with motivation and attitude.

The reliability test for Factor 2 was more involved. An initial reliability test for the six items that loaded onto Factor 2 provided a Cronbach's Alpha of 0.70, just at the generally accepted cut-off of 0.70 (Pallant, 2010). The item-total correlation value for the item "improvement by initially low performing student" was 0.29. This item didn't seem to fit with the other items on this scale conceptually and had a low item to total correlation value, so following Pallant's (2010) advice this item was removed. A subsequent measure of Cronbach's alpha was undertaken with this item removed, producing a value of 0.70. However, Examination of the Item-Total Statistics showed that the item "assessments of recall or understanding" had a low Corrected Item-Total Correlation of 0.32. An examination of this scale with the structural components previously identified in the research provides support for removing this second item as well. Anecdotally, my experiences in the B Ed. program and my experiences working in schools enforced the notion that assessments of recall are lower quality assessments and are contrasted with more involved assessments such as projects, authentic assessments, or

more intellectually demanding assessments such as reasoning, higher order thinking or mastery of learning objectives. As a result of this analysis, I reduced the items on this scale from six to four items, without impacting the internal consistency. This final scale resembled the factor “higher order thinking” identified by Duncan and Noonan (2007) as being the sole factor for the structural component “Cognitive Levels of Assessment,

Factor 3 was initially comprised of 6 items with an internal consistency coefficient of 0.66. A Corrected Item-Total Correlation value for ‘major exams or compositions’ was 0.21 and an adjusted Cronbach’s alpha of 0.67 if the item was deleted, so this item was removed from the scale. This reliability coefficient is still too low to meet the cut-off of acceptability at 0.70. However, the items on this scale: “compared to previous students,” “compared to students that semester,” “set scale of percentage correct,” “performance on quizzes,” and “how the student did on homework” do compare favourably to the second factor identified by Duncan and Noonan (2007) as falling under the structural component “grading practices: use of external benchmarks.”

Factor 4 with its original six items generated a reliability coefficient of 0.75. However, the Inter-Item Correlation Matrix indicated that the item “objective assessments (e.g. multiple choice, matching, short answer)” had correlations of .3 or less with all other items on the scale. Again, this suggests that this item is measuring something different than the rest of the items on the scale. The removal of this item improved Cronbach’s alpha to 0.76. This decision was supported by a logical analysis of the items on this scale. All the other items dealt with assessments of a constructed response format, the antithesis of objective assessments like multiple choice. The final version of this scale was composed of 5 items and corresponded to the category of

“constructed response assessments,” one of two factors found under the structural component of “assessment strategies” by Duncan and Noonan (2007)

The P.C.A. also identified a fifth and final factor comprised of 6 items. The reliability coefficient of these 6 items was 0.69. One item, “including zeros” showed very low values (all below .30) on the Inter-Item Correlation Matrix. Acknowledging what previous research had to say about this item proving problematic, it was removed from the scale and the reliability coefficient increased to 0.74. The majority of items on this scale dealt with effort save,” improvement in performance” and “overall ability level of the student.” This factor poses a challenge to interpretation using Duncan and Noonan’s (2007) factor structure. Logically, items dealing with effort should fall into the factor of “academic enabling behaviors” with other items in Factor 1. Why beginning teachers’ view issues of student effort as a separate category from motivation and attitude is interesting. The literature review previously discussed the impact that motivation can have on students’ effort and performance depending on the students’ view of intelligence (Dweck, 1986; 2010). Further to this, Brookhart (2007) argues that feedback from teachers to students about effort, instead of outcomes, is unhelpful in improving student performance.

Table 18 presents the final internal consistency results and logical analysis on the items on the 5-factor scales across the three structural components of the third section of the survey’s internal structure. These results do not conform to those generated by Duncan and Noonan (2007). They had only two factors under grading practices, whereas my results found three factors (with two sub categories under academic enabling behaviors: motivation and attitude, and effort and ability). Duncan and Noonan (2007)

also identified two factors under the assessment strategies construct: constructed response items and grouped quizzes/grouped quizzes. My study found results for constructed response items only. Both my study and Duncan and Noonan (2007) both found a factor for higher order thinking.

Table 18

*Reliability Coefficients for Factor Structure of Assessment Practice Section*

Structural Component	Factor	No. of Items	Alpha
Grading Practices		20	0.89
	Motivation and attitude	10	0.89
	Effort and Ability	5	0.74
	Use of external benchmarks	5	0.67
Assessment Strategies		5	0.76
	Constructed response assessments	5	0.76
Cognitive Levels of Assessment		4	
	Higher order thinking	4	0.70

Tables 19-23 provide the descriptive statistics for beginning teachers' responses to the assessment practice portion of the survey after the original 35 items had been reorganized into 5 scales with problematic items removed. The scores on the individual items in each factor were then averaged together to create an overall score for that scale. The lowest average score was found for the structural component of grading practices: use of external benchmarks ( $M=2.84$ ,  $SD= 0.69$ ). This reflects an average response between "Very little" and just short of "some," suggesting that these candidates will not commonly use these benchmarks. Many items on this scale dealt with making comparisons between students, a form of norm-referenced assessment and a practice

frowned upon by the research literature and by *Growing Success* (O.M.E., 2010). Such a low score on this scale suggests that the teacher candidates who responded to this survey do not believe they will make norm-based assessments the majority of their classroom assessment work. As will be discussed later in the Inferential Statistics section, there were no statistically significant differences between groups, hence my focus on only the overall group statistic.

The second lowest average score was also found under the structural component of grading practices: motivation and attitude ( $M=3.18$ ,  $SD= 0.75$ ). There no statistically significant differences between groups (see Inferential Statistics section) so here I treat only with the overall group score. This result is mixed, with respondents stating that they would allow academic enabling behaviours (which are irrelevant to whether the student has performed adequately on the subject matter/skills under investigation) to influence their assessment decisions somewhat. This assessment practice is not supported in the research literature (Black & William, 1998a; Young & Kim, 2010). However, the importance of motivation has been demonstrated in the literature review (in particular Dweck, 1986, 2010) and it is possible that teacher candidates who plan to reward motivation are not wholly in the wrong. The structural component of assessment strategies comes next with use of constructed response formats ( $M=3.41$ ,  $SD= .81$ ). *Growing Success* (O.M.E., 2010) stresses that teachers are to use a variety of assessment formats in their practice, many of which are included on the items for this scale. This factor did demonstrate a statistically significant difference between groups (Grade level: 7&8 versus 11&12). This finding is discussed in greater detail in the Inferential Statistics section.

The final factor under the structural component of grading practices was effort and ability ( $M=3.95$ ,  $SD= 0.75$ ). There were no statistically significant differences between groups. This scale had the highest average overall group score, and the teacher candidates generally believed they would use these items “somewhat” to “quite a bit.” Such a result is problematic from a criterion-referenced approach to assessment since the criteria under assessment generally deal with academic or content related areas and not the personal characteristics of students. However, just as in the case of motivation and attitude above, such considerations on the part of teacher candidates’ in their assessment practice does find support in the work of Dweck (1986, 2010).

The highest scale score was for the structural component “cognitive level of assessment” and the factor was “higher order thinking” ( $M=4.04$ ,  $SD=0.68$ ). A score spreading from the top end of “somewhat” to the low end of “quite a bit” is a hopeful sign that beginning teachers’ assessment practices reflect the importance placed upon 21<sup>st</sup> century skills found in the research literature and in *Growing Success* (O.M.E., 2010).

Table 19

*Factor 1 of Assessment Practice: Motivation and Attitude*

Group	Sub-group	Mean (SE)	SD	Skewness (SE)	Kurtosis (SE)
Grade	7 & 8	3.76 (0.28)	0.75	-1.90 (0.79)	3.55 (1.59)
	9 & 10	3.19 (0.08)	0.77	0.23 (0.25)	0.32 (0.49)
	11 & 12	3.12 (0.09)	0.69	0.14 (0.32)	-0.99 (0.63)
Subject	M & S	3.07 (0.10)	0.75	0.42(0.31)	-0.42 (0.61)
	L/H/SS	3.15 (0.10)	0.72	0.13 (0.31)	-0.15 (0.61)
	A/PE/T	3.41 (0.13)	0.76	-0.30 (0.39)	1.25 (0.77)
Level	Academic	3.11 (0.08)	0.71	-0.34 (0.26)	3.17 (0.51)
	Applied	3.17 (0.14)	0.79	0.39 (0.40)	0.66 (0.78)
	Other	3.37 (0.14)	0.78	-0.73 (0.40)	0.43 (0.78)
Program	Concurrent	3.17 (0.10)	0.74	0.60 (0.30)	3.20 (0.60)
	Consecutive	3.20 (0.08)	0.76	0.14 (0.25)	-0.62 (0.49)
Overall		3.18 (0.06)	0.75	0.15 (0.19)	-0.21 (0.39)

Table 20

*Factor 2 of Assessment Practice: Higher Order Thinking*

Group	Sub-group	Mean (SE)	SD	Skewness (SE)	Kurtosis (SE)
Grade	7 & 8	4.50 (0.21)	0.56	-0.63 (0.79)	1.21 (1.59)
	9 & 10	3.97 (0.07)	0.69	0.34 (0.25)	0.07 (0.49)
	11 & 12	4.08 (0.09)	0.64	0.22 (0.32)	0.24 (0.63)
Subject	M & S	4.07 (0.09)	0.66	0.45(0.31)	0.62 (0.61)
	L/H/SS	4.00 (0.09)	0.68	0.28 (0.31)	-0.01 (0.61)
	A/PE/T	4.05 (0.12)	0.72	-0.12 (0.39)	-0.64 (0.77)
Level	Academic	4.09 (0.07)	0.67	0.50 (0.26)	0.37 (0.51)
	Applied	3.80 (0.12)	0.71	0.20 (0.40)	-0.27 (0.78)
	Other	4.16 (0.11)	0.61	-0.23 (0.40)	-0.97 (0.78)
Program	Concurrent	4.06 (0.09)	0.70	0.25 (0.30)	0.30 (0.60)
	Consecutive	4.02 (0.07)	0.67	0.21 (0.25)	-0.22 (0.49)
Overall		4.04 (0.05)	0.68	0.23 (0.19)	-0.03 (0.39)

Table 21

*Factor 3 of Assessment Practice: External Benchmarks*

Group	Sub-group	Mean (SE)	SD	Skewness (SE)	Kurtosis (SE)
Grade	7 & 8	3.03 (0.13)	0.35	-0.30(0.79)	2.38 (1.59)
	9 & 10	2.85 (0.07)	0.72	0.41(0.25)	0.04 (0.49)
	11 & 12	2.80 (0.09)	0.68	0.16 (0.32)	-0.15 (0.63)
Subject	M & S	2.90 (0.09)	0.70	0.37(0.31)	0.29 (0.61)
	L/H/SS	2.74 (0.08)	0.60	0.14 (0.31)	0.19 (0.61)
	A/PE/T	2.90 (0.14)	0.81	0.16 (0.39)	-0.66 (0.77)
Level	Academic	2.80 (0.08)	0.72	0.44 (0.26)	0.12 (0.51)
	Applied	3.03 (0.12)	0.72	0.03 (0.40)	-0.12 (0.78)
	Other	2.75 (0.09)	0.55	-0.18 (0.40)	-0.30 (0.78)
Program	Concurrent	2.76 (0.08)	0.65	0.34 (0.30)	0.28 (0.60)
	Consecutive	2.89 (0.07)	0.72	0.25 (0.25)	-0.07 (0.49)
Overall		2.84 (0.06)	0.69	0.30 (0.19)	0.02 (0.39)

Table 22

*Factor 4 of Assessment Practice: Constructed Response*

Group	Sub-group	Mean (SE)	SD	Skewness (SE)	Kurtosis (SE)
Grade	7 & 8	4.05 (0.34)	0.91	-0.65 (0.79)	-0.95 (1.59)
	9 & 10	3.40 (0.09)	0.82	0.38 (0.25)	0.56 (0.49)
	11 & 12	3.30 (0.10)	0.76	-0.30 (0.32)	-0.52 (0.63)
Subject	M & S	3.07 (0.12)	0.91	0.57(0.31)	-0.14 (0.61)
	L/H/SS	3.61 (0.08)	0.64	1.08 (0.31)	0.65 (0.61)
	A/PE/T	3.62 (0.13)	0.72	-0.28 (0.39)	0.56 (0.77)
Level	Academic	3.33 (0.09)	0.84	0.36 (0.26)	0.47 (0.51)
	Applied	3.30 (0.11)	0.63	0.19 (0.40)	0.43 (0.78)
	Other	3.66 (0.16)	0.89	-0.51 (0.40)	-0.11 (0.78)
Program	Concurrent	3.21 (0.12)	0.91	0.42 (0.30)	0.49 (0.60)
	Consecutive	3.52 (0.08)	0.71	0.20 (0.25)	-0.41 (0.49)
Overall		3.40 (0.07)	0.81	0.17 (0.19)	0.35 (0.39)

Table 23

*Factor 5 of Assessment Practice: Effort and Ability*

Group	Sub-group	Mean (SE)	SD	Skewness (SE)	Kurtosis (SE)
Grade	7 & 8	4.31 (0.31)	0.82	-1.96 (0.79)	4.04 (1.59)
	9 & 10	3.92 (0.08)	0.80	-0.17 (0.25)	0.51 (0.49)
	11 & 12	3.95 (0.09)	0.66	-0.33 (0.32)	-0.02 (0.63)
Subject	M & S	3.91 (0.10)	0.72	-0.73(0.31)	0.22 (0.61)
	L/H/SS	3.89 (0.07)	0.77	0.07 (0.31)	0.82 (0.61)
	A/PE/T	4.12 (0.13)	0.77	-0.31 (0.39)	-0.64 (0.77)
Level	Academic	3.98 (0.08)	0.69	0.03 (0.26)	-0.20 (0.51)
	Applied	3.70 (0.13)	0.78	-0.51 (0.40)	1.04 (0.78)
	Other	4.15 (0.14)	0.80	-0.60 (0.40)	0.11 (0.78)
Program	Concurrent	3.97 (0.10)	0.76	-0.40 (0.30)	1.38 (0.60)
	Consecutive	3.93 (0.08)	0.75	-0.18 (0.25)	-0.64 (0.49)
Overall		3.95 (0.06)	0.75	-0.27 (0.19)	0.35 (0.39)

**Inferential Statistics**

**Assessment Purposes.** The descriptive statistics for these questions are provided in Tables 2-6. The highest score of the five questions dealing with assessment purpose was for “Providing Feedback” ( $M= 4.44$ ,  $SD=0.67$ ) followed by similar questions dealing with different types of formative assessment: “Majority of Assessments will be for Formative Purposes” ( $M= 4.08$ ,  $SD= 0.81$ ) and “Assess Primarily to inform my teaching” ( $M= 3.92$ ,  $SD= 0.71$ ). Two questions regarding summative assessments had the lowest scores: “Majority of assessments will be for Summative Purposes” ( $M= 3.55$ ,  $SD= 0.74$ ) and “Assess Primarily to Assign Grades” ( $M= 3.46$ ,  $SD= 0.84$ ). I conducted a paired samples  $t$ -test to determine if the scores awarded to each of the five questions regarding assessment purposes were significantly different ( $\alpha = 0.05$ ). Only one pair of items

failed to record a significant difference in scores; “7a assess primarily to assign grades” ( $M= 3.46, S.D.=0.84$ ) and 8a “the majority of my assessments will be for summative purposes” ( $M= 3.55, S.D.= .74$ ) These results suggest that teacher candidates prefer the use of assessments for formative purposes.

The results of a 3 X 3 X 3 X 2 (Grade X Subject X Academic level X Program stream) ANOVA indicated no significant differences between these groups for teachers’ proposed assessment purposes. Two interaction effects were significant at the initial 0.05 level (Grade X Subject and Grade X Subject X Level), but the Levene’s Test of Equality of Variance indicated that the assumption of equality of variances had been violated ( $p=0.012$ ). In such an instance, both Pallant (2010) and Tabachnik and Fidell (2007) recommend increasing the stringency of the significance level to 0.01. At this level of significance, the interaction effects previously identified were no longer statistically significant.

**Assessment Practice.** The descriptive statistics for the five scales generated from the 35 item section of teacher candidates’ summative assessment practice are found in Tables 19-23. A paired samples t-test on all possible pairs was conducted to determine if the scores for the five scales were significantly different ( $\alpha= 0.05$ ). All pairs were statistically significantly different at the 0.01 level save one: higher order thinking ( $M= 4.04, SD=0.68$ ) & effort and ability ( $M=3.95, SD= 0.75$ ). The results of the five 3 X 3 X 3 X 2 (Grade X Subject X Academic level X Program stream) ANOVAs indicated no significant differences ( $\alpha= 0.05$ ) between these groups on their proposed assessment for four of the scales: motivation and attitude, higher order thinking, use of external benchmarks, and effort. In contrast, Duncan and Noonan (2007) found significant main

effects for subject area on four of their five scales. I did find a statistically significant difference between two groups on the constructed response scale. The teacher candidates identified as teaching a Grade 7 or Grade 8 class ( $M= 4.05, S.D.= 0.9$ ) scored significantly higher on this scale than the teacher candidates who indicated they were teaching Grades 11 and 12 ( $M=3.30, S.D.= 0.76$ ). A partial eta squared value of 0.79 meant that almost eight percent of the variance in scores on this question could be accounted for by program stream. To transform the partial eta squared value into Cohen's  $d$ , I subtracted the Grade 7&8 mean score from the Grade 11&12 mean score and divided this value by their pooled standard deviation. The result was  $d= 0.9$ , a large effect size. It appears that teacher candidates hoping to teach in the intermediate grades intend to put extra emphasis on this particular assessment format for students at this level, whereas teacher candidates hoping to teach in the senior grades will not.

**Values and Beliefs.** This section asked teacher candidates to rate the importance of the four values to their proposed assessment practice. Tables 7-11 show the descriptive statistics for these scores. A paired samples  $t$ -test was conducted on the overall teacher candidates' scores on each pair of the four values to determine if the differences in mean scores for the four values were statistically significant ( $\alpha=0.05$ ). The highest scoring value, Equity, was significantly different from the other four values ( $M=4.64, SD=0.52$ ). The lowest scoring value, Efficiency, was also significantly lower from the other values ( $M=3.06, SD= 0.88$ ). The remaining two values (Choice, Excellence) were not statistically significantly different from one another. These results are consistent with the literature regarding the importance of efficiency and routinization of teachers' assessment practices and the importance given to the value of Equity in *Growing Success* (O.M.E.,

2010). Next, four 3 X 3 X 3 X 2 (Grade X Subject X Academic level X Program stream) ANOVAs were conducted to determine if there were significant between group differences on the four dependent measures of beliefs and values. No significant differences ( $\alpha=0.05$ ) were found across the four dependent measures.

The questions in the rest of this section of the survey required teacher candidates to rank the four values in order of importance to their assessment practices. The responses to these questions were problematic. Many respondents failed to rank the values in order of importance (instead merely repeating the process they undertook in the previous portion of this section), inverted the scale, or left it blank. Consequently, the responses from this portion could not be used reliably, so these responses were not reported, and were not used in any subsequent inferential analyses.

**Influence of Experiences and Beliefs on Assessment.** The descriptive statistics for these questions are found in Tables 14-16. Paired samples *t*-tests on all possible pairs were conducted on each possible pair to determine if the scores for the three questions regarding experiences and beliefs were significantly different from one another ( $\alpha=0.05$ ). All pairs were significantly different. Next, three 3 X 3 X 3 X 2 (Grade X Subject X Academic level X Program stream) ANOVAs were conducted ( $\alpha=0.05$ ). There were no statistically significant differences between groups regarding the influence of their beliefs before taking the B.Ed on their proposed future assessment practices. Nor was there a statistically significant difference between groups in their esteem for the influence of their practicum on their proposed future assessment practices. Interestingly, for B. Ed. course work, there was an interaction effect significant at the initial 0.05 level (Grade X Subject), but a significant Levene's Test of Equality of Variance indicated that

the assumption of equality of variances had been violated ( $p=0.02$ ). Increasing the stringency of the significance level to 0.01 resulted in no significant interaction effects. There was still a main effect ( $p=.001$ ) between program types: concurrent ( $M=3.03$ ,  $SD=1.04$ ) and consecutive ( $M=3.52$ ,  $SD=.95$ ). A partial eta squared value of .52 meant that just slightly more than five percent of the variance in scores on this question could be accounted for by program stream. The effect size for this difference was  $d= 0.49$ . This result indicates a close to medium effect size. This is an intriguing result considering that concurrent candidates undergo 4 years of pedagogical course work as opposed to the one year undertaken by consecutive candidates. Such a result might perhaps be a result of the perception by concurrent candidate of course work's diminishing utility across the four years. This hypothesis would suggest that concurrent candidates value their class work less than consecutive candidates, but other hypotheses are available.

## **Chapter 5**

### **Discussion**

In order to better understand the assessment practices of beginning teachers and how these might align with the supported or recommended practices outlined in the research and *Growing Success* (O.M.E., 2010) I examined teacher candidates' self-reports in order to address the three research questions: (1) What do teacher candidates imagine will be the primary assessment purposes, assessment practices and assessment formats in their classroom? (2) What are teacher candidates' values and beliefs surrounding assessment? and (3) What contextual factors influence teacher candidates' assessment beliefs and practices?

#### **Limitations and Difficulties in Interpreting Results**

Given the nature of my research, using a convenience sample of teacher candidates who volunteered to complete the survey, it is important to acknowledge the limitations in the research. The research design did not allow the possibility of a truly random sample, and this impinges on the generalizability of my findings. Furthermore, the sample used in my research and the larger population of intermediate and senior candidates from which it was drawn may not be indicative of teacher candidates elsewhere in the province.

Additionally, there is always a difficulty involved in interpreting self-reported responses. I cannot be certain that teacher candidates interpreted the definitions or terms in my survey in the way I intended. For example, teacher candidates may have interpreted the values questions using their own definitions for the four values. Self-reporting can sometimes fail to be accurate if respondents, such as the teacher candidates who

answered my survey, feel pressured to answer in a particular way. This could very well have been the case regarding the low values teacher candidates ranked the value of efficiency. Teacher candidates could have felt a certain amount of pressure to rank this value lower than other more socially acceptable or promoted values such as equity and excellence.

Furthermore, it is possible that teacher candidates used a more formal definition for assessment when answering survey questions. If teacher candidates answered the survey using a definition for assessment more closely related to Yin, et al. (2008) “formal embedded assessments” and not the two other types of assessment “on-the-fly” and “planned-for-interaction,” it’s possible that valuable information about teacher candidates classroom assessment practices may have been missed.

Finally, the information provided by teacher candidates regarding their values and beliefs, assessment purposes, assessment formats and summative assessment practices may not reflect the reality they face when they become full time classroom teachers. It may be that the positive alignment found between my sample’s responses, and the policy and research literature disappears once these teacher candidates are faced with the task of running their own classrooms. Certainly, the research covered in my review indicates classroom teachers engage in poor assessment practice, and place too much of an emphasis on grading and classroom management.

### **Assessment Purpose**

Keeping in mind the caveat above, about teacher candidates’ definition of assessment and their conception of what assessment looks like (formal, vs on-the-fly) teacher candidates’ responses to the assessment purpose section of the survey suggest that

they were aware of the distinction between assessments based on purpose. Teacher candidates' stated beliefs that the majority of their future assessments will be used for formative purposes conform to the current research literature regarding the primary focus of teacher assessments (Black & Wiliam, 1998a, 1998b, 1998c; Brookhart, 2007). In particular, the teacher candidates' beliefs that the primary purpose of their assessment practices will be to provide feedback to their students is a positive finding.

The results of my research regarding teacher candidates' beliefs about the primary purposes of assessment are also congruent with *Growing Success* (O.M.E., 2010). *Growing Success* stresses the importance of formative assessment and the role of descriptive feedback and the connection these practices have with developing students' self-assessment skills. Again, seeing the compatibility of teacher candidates' beliefs about the importance of Assessment *for* and Assessment *as* learning, and the emphasis placed on these forms of assessment in *Growing Success*, suggests that a certain common sense to assessment is shared between teachers and policy makers. A shared commonsense or 'habitus' between beginning teachers, researchers and policy makers may mean that the discord identified by Blackmore (2010) is not as severe as thought. This shared understanding removes one possible source of misalignment between the expectations created by the policy environment and the actual assessment practices of teachers. It remains to be seen whether this shared habitus will remain. It's possible that the beginning teachers who responded to my survey do privilege the doxa of researchers and policy makers more now than they will as their careers continue.

## Assessment Practice

Even given the potential difficulties inherent in using self-reported data, the results from my study in the areas of values and beliefs, assessment purpose, and assessment format largely reflect supported practices in the literature, and the results from the summative assessment practice section continue this trend. Teacher candidates' scores on the 5-item third factor "Use of External Benchmarks" were relatively low ( $M=2.35$ ,  $SD = 1.08$  and Overall  $M= 2.84$ ,  $SD = 0.69$ ) corresponding to a usage of "very little" on the 6-point Likert scale. In particular, scores on the questions as to whether or not they would compare students to one another in the same semester and across semesters were also low (Overall  $M=2.35$ ,  $SD = 1.08$  and Overall  $M= 2.07$ ,  $SD = 1.11$ ) reflecting a usage of "very little." This suggests that these teacher candidates may be less influenced by the formerly used norm-referenced framework for making their assessment judgments.

While research indicates that the questions teachers ask of students are of limited quality in assessing students' understanding (Black & Wiliam, 1998a; Stiggins, 1988; Young & Kim, 2010), the second factor in the assessment practices section of the survey was "Higher order thinking." This factor contained four items in the final analysis and its rating (Overall  $M= 4.04$ ,  $SD = .68$ ) corresponds to "quite a bit" on the 6-point Likert scale. This finding perhaps should be tempered with results from the fourth factor identified in my P.C.A. the use of "Constructed Response" assessment formats (overall  $M= 3.40$ ,  $SD = .80$ ) which found that teacher candidates rated their use of these at a level corresponding to "somewhat" on the 6-point scale used for the five items in this factor. Given that constructed response question formats are often held up as being capable of assessing students' understandings more deeply (see Torrance & Pryor, 2001 and their

discussion of convergent and divergent questions) and thus elicit higher order thinking, it remains an open question how committed the beginning teachers are to eliciting evidence of higher order thinking from their future students. It may be the case that while teacher candidates' are committed to eliciting higher order thinking from their students, but they feel that time constraints and other considerations might inhibit the frequency at which they employ constructed response assessments. This hypothesis has interesting implications for the results regarding the relative importance teacher candidates place on the value of efficiency. It may be the case that the results about the frequency teacher candidates plan to use different assessment formats are indicative of a concern for efficiency that wasn't clearly brought out in the values questions. In addition to exploring this possibility further, more research into teacher candidates' understandings of what qualifies as "Higher Order Thinking" would be helpful, as would further exploration of teacher candidates' beliefs about the feasibility of a high frequency of constructed response assessment formats. Further research would also likely include examination of the examples of the types of questions teachers generate for "Constructed Response" assessments to determine if they do in fact target "Higher Order Thinking."

Perhaps most troubling however, were the findings regarding the last two scales within the summative assessment practice section of the survey. These two factors dealt with students' behaviours and attitudes. Factor 1, "motivation and attitude" contained 10 items and had an average score associated with being used "somewhat" on the 6-point Likert scale (overall  $M= 3.18$ ,  $SD = 0.75$ ). Factor 5, "effort and ability," with its 5 items, had an even higher average rating, just below the category "quite a bit." These findings suggest that teacher candidates intend to consider the non-academic enabling behaviours

of students in making their summative assessment judgments (overall  $M= 3.95$ ,  $SD = 0.75$ ). Such use of non-achievement related measures is considered problematic if it is included as part of a grade in the research literature (e.g., Cizek, et al., 1995/1996).

However, while the inclusion of non-achievement measures in the determination of students' final grades is a cause of criticism in some parts of the research literature (particularly as it relates to the validity of teachers measurement of student performance), other research indicates that the inclusion of such measures may be warranted. Dweck's (1986, 2010) research indicates that students' non-cognitive abilities influence their performance in school. Teachers who believe in the incremental theory of intelligence, who value the improvements made by students who were at first low performers, and who reward persistence and effort in the face of challenge may influence students to share these same views and cause those students to persevere and perform better than they have otherwise. In particular, the average scores for question 9c (see Appendix B) "improved performance" (overall  $M= 4.55$ ,  $SD = 0.99$ ) and question 9p "improved performance of initially low performing student" (overall  $M= 4.08$ ,  $SD = 4.08$ ) both rated by teachers as something they would let influence their summative assessment practices "quite a bit" suggest that teachers are ready to reward those students who work hard to improve. If such is true then the average scores obtained for Factor 1 "motivation and attitude" and Factor 2 "effort and ability" might actually prove beneficial to those students who most need help in achieving their learning goals.

The results from the assessment practice section of the survey track closely to what is asked of teachers' assessment practices in the *Growing Success* (O.M.E., 2010) document. Teacher candidates' disinclination to use norm referencing in their

assessments reflects the document's focus on criteria referenced assessment. The relative importance placed on higher order thinking in teacher candidates' responses is also encouraged in *Growing Success*. *Growing Success* recommends the use of many different assessment formats that allow students to demonstrate the full depth of their learning, an approach that finds some support in the responses of teacher candidates that they would use constructed response formats "somewhat" for their assessments.

Any misalignment between *Growing Success* (O.M.E., 2010) and the teacher candidates' approaches to summative assessment lies in their proposed use of non-achievement related measures. According to *Growing Success*, teachers should only consider achievement measures in the generation of students' grades. However, there is also an expectation that teachers will reinforce the positive learning skills and work habits it considers important for later in life. Teachers can certainly report on these skills and work habits separately on report cards, but it is possible that teachers believe the only way to ensure students will value these skills is to include these behaviours in the calculation of final grades. It is also possible that teachers make these decisions regarding the use of non-achievement related criteria less deliberately. It might be the case that these decisions are being made based on unexamined, tacit assumptions held by teachers that reflect underlying values and beliefs or their previous experiences as students.

### **Values and Beliefs**

Teacher candidates' responses to the values and beliefs questions on the survey provide evidence there is alignment between the values and beliefs teachers hold and the values and beliefs implicit in the research literature and those espoused by the *Growing Success* (O.M.E, 2010) document. Further, the teacher candidates' focus on equity as the

value most important to them shows a concern for all learners, not just those identified as the best or the strongest, a position in agreement with Shepard's (2000) description of the current approach to teaching and learning. The focus on equity also eclipses teacher candidates' concerns regarding their own autonomy and the choices they make in their classrooms, identified by Fuller (1969) as the fourth level of teacher development.

If we take teacher candidates' responses at face value, and leave aside for the moment concerns about their interpretations of the different values or the possible pressures they felt in rating their importance, the candidates' lack of concerns with the value of efficiency is potentially troubling. This is not to say that a focus on the value of equity as being the most important value informing their teaching is wrong. Rather, as identified in the research by Berliner (2004), expert teachers (those most effective in the classroom in improving student learning) rely on habits of routinization and automaticity in their classroom practice:

Expert teachers often develop automaticity and routinization for the repetitive operations that are needed to accomplish their goals... expert teachers have fast and accurate pattern-recognition capabilities, whereas novices cannot always make sense of what they experience; expert teachers perceive meaningful patterns in the domain in which they are experienced; and although expert teachers may begin to solve problems slower, they bring richer and more personal sources of information to bear on the problem that they are trying to solve. (Berliner, 2004, p200)

Expert teachers who use automaticity and routinization in their classroom practice are likely to avoid some of the problems involving classroom management. In addition, rather than merely streamlining classroom interactions (to the detriment of student

learning as referenced in the literature see Shavelson & Stern, 1981), expert teachers are able to do more. Expert teachers who have built up a storehouse of patterns for interpreting students' responses to assessments are much more likely to use this information to design assessments that provide the information they are looking for (degree of student understanding) and avoid the features of assessments that provide irrelevant or distracting information. Such focused and targeted assessments mean that teachers can spend more time providing specific feedback to each of their students. Efficiency then is a means of achieving the goal of equity, a necessary but not sufficient condition for improving student learning. Expert teachers who use their experience to provide feedback and learning opportunities to specific students based on those students needs are much more likely to improve the learning outcomes for all their students

In addition, effective assessments that target students individually require an investment of time on the part of teachers. Doyle and Ponder (1977) note that cost is one of the three main factors teachers consider when determining the practicality of adopting an educational practice. The provision of student specific feedback from assessments can certainly be time consuming. The teacher candidates who completed my survey were about to begin their teaching careers. It's likely that as they progress through their careers they will begin to realize the importance of efficiency to achieving their assessment goals, founded as they are on the value of Equity.

The teacher candidates' focus on Equity also aligns nicely with the *Growing Success* document: "Every student is unique and each must have opportunities to achieve success according to his or her own interests, abilities and goals" (O.M.E., 2010, p. 1). Such an alignment between respondents' values and those underlying the approach to

assessment articulated in the policy document is heartening for those hoping that teachers will enact the approaches *Growing Success* (O.M.E., 2010) recommends.

### **Assessment Format**

Teacher candidates' commitment to equity as demonstrated in the values and beliefs section suggest that teacher candidates' commitments to equity are recapitulated in the examples of assessments they intended to provide (both formative and summative). Many responses made explicit mention of fairness and giving students a choice in how they could demonstrate their learning. The recurring mention of Assessment *as* learning in the assessment purpose and rationale section of the formative assessment format section, as well as the frequency of 'feedback' as an answer indicate that respondents' choices of assessment formats cohere with both their values regarding assessment practices, as well as their primary assessment purposes. The concern teacher candidates had with ensuring that their summative assessments be comprehensive through the use of multiple question types also finds support in the work of Torrance and Pryor (2001) and their distinction between convergent and divergent questions. This alignment between values, purpose and format is also found in what the research literature has identified as recommended practices (e.g., Black & William, 1998a, Brookhart 2007, Hattie & Timperley, 2007).

The varied assessment formats identified by teacher candidates adhere to the approach recommended in *Growing Success*, where teachers are asked to provide assessments that are "ongoing, varied in nature" (O.M.E., 2010, p. 6). Additionally, the focus on equity and fairness in teacher candidates' justifications for their assessment formats are, as has already been shown, reflective of the practices asked of teachers in

*Growing Success*. Teacher candidates' assessment format choices that highlighted student reflection "tickets out of class" and "reflective/self-assessments" targets those learning skills and work habits identified in *Growing Success* as being necessary in school and life (and are the same skills identified by Darling-Hammond, 2005).

### **Influence of Experiences and Beliefs on Assessment**

Teacher candidates identified their practicum experiences as being the most influential factor on their proposed future assessment practices ("quite a bit" overall  $M = 4.6$ ,  $SD = 0.66$ ). This finding corresponds to the research done by Doyle and Ponder (1977) that provides a model for understanding teachers as "pragmatic skeptics." Their model, identifying the need for teachers to feel that the information or techniques they are being provided are "practical", helps explain why teacher candidates view their practicum experiences so highly. The concern of course is that teacher candidates may allow their experiences during their practicum to override what they were taught as being sound practice during their B. Ed. course work. Given the extent to which the research literature identifies problems with practicing teachers' approaches to education, this concern is warranted.

Teacher candidates considered their B. Ed. course work to have influenced them "somewhat" (overall  $M = 3.33$ ,  $SD = 0.08$ ), providing some solace to those concerned that teacher candidates who experience poor assessment practices in their practicum are doomed to emulate it. In addition, teacher candidates ranked the B. Ed. course work somewhat more influential than the beliefs they held before entering the B. Ed. Program. Hence there is evidence that their time on campus in B. Ed. courses is useful to their development as teachers.

## Differences between Groups

There were no significant differences between the groups of teacher candidates on the sections of my survey dealing with “values and beliefs,” and “assessment purpose.” There was a statistically significant difference between groups on the “constructed response” scale of the summative assessment practice section. Grades 7 and 8 teacher candidates were more likely to use constructed response items than those who had been most recently teaching Grades 11 and 12. The small sample size for the candidates who had been teaching Grades 7 and 8 ( $n=7$ ) was an issue, but the results are consistent with other research (MacMillan, 2001; Young & Kim, 2010). These candidates appear to have a greater desire to use assessments comprising an oral component (with more time during a whole day of instruction and interaction available to them) than in High School classes (operating on a standard 75 minute class period in most cases). It might also be the case that High School teacher candidates feel they have more learning objectives to cover in their classes and objective assessments are a way of covering more material quicker.

The only other statistically significant main effect between groups was for the question dealing with the influence of B. Ed. course work on teacher candidates’ future assessment practices. Teacher candidates in the concurrent program rated their course work as less influential ( $M= 3.03$ ,  $S.D. =1.04$ ) than teacher candidates in the consecutive program ( $M= 3.52$ ,  $S.D.= .95$ ). This difference reflects a medium effect size ( $d=0.49$ ). One explanation is that concurrent program teacher candidates experience diminishing utility of their course work over their five years in the program.

Training and certification is important for teacher efficacy (Berliner, 2001; Darling-Hammond, 2000a, 2000b; Darling-Hammond, Chung & Frelow, 2002; Darling-

Hammond, 2005; Darling-Hammond, 2009; Ferguson, 1991). However, the majority of these studies compare teachers who have undergone some form of comprehensive training to teachers who received very little if any such training. The results of my study seem to indicate that there is no difference between teachers who receive five years of pedagogical training versus those who only receive one in the areas of the two groups values and beliefs, assessment purposes, use of different assessment formats, or summative assessment practices. One potential explanation for this is that teacher expertise is something that takes several years to develop. Berliner's (2004) research indicates that it takes a minimum of 3 to 5 years *in schools* before such expertise develops. It could be that it is the "in school" experience that separates expert teachers from non-experts. Alternatively, it could be that my results, based as they are on teacher candidates just about to start their careers, come too early to notice effectively the differences between groups. It might be the case that concurrent candidates mature into expert teachers more quickly than consecutive candidates once they have started full time in the classroom. Further, the assessment training and the majority of practicum teaching experience occur in the final year of the B.Ed program, a time that is shared by concurrent and consecutive teacher candidates.

### **Final Thoughts**

Perhaps my results can be seen as providing a baseline reading of teacher candidates' values and beliefs, assessment purposes, assessment formats and summative assessment practice. Further research, examining whether these teacher candidates continue to display the habits once they are in the classroom for an appreciable period of time may shed light on whether the realities of classroom life change teacher values,

beliefs and practices surrounding assessment and whether this change demonstrates alignment with what previous research has found.

It is likely that the lack of appreciable classroom experience by teacher candidates that led to the presence of very few differences between groups in my analyses. Unlike Duncan and Noonan (2007), I was unable to distinguish between groups regarding their different assessment practices. Teacher candidates' lack of experiences in context specific classrooms (grade, subject, academic level) made for very homogenous responses.

What is clear is that the teacher candidates who completed the survey described values and beliefs, and assessment practices in line with the recommendations from the research and policy literature. These teacher candidates did privilege the values of equity and fairness when dealing with students. They professed to want to make formative assessment, particularly assessment *as* learning, a priority in their classrooms. Assessment formats that relied upon feedback and reflection were frequently mentioned as being employed by these future teachers. While their summative assessment practices may have proved problematic from an achievement criteria perspective, they are supported by the literature surrounding improving student motivation, and increasing persistence and challenge seeking.

These findings suggest that teacher candidates are willing to engage in assessment practices that promote the skills and work habits valued by researchers and policy makers. My results indicate that teacher candidates are concerned with equity for all their students, are interested in using formative assessment to improve their instruction and student learning outcomes, privilege assessment formats that allow students to demonstrate their learning in myriad ways that reflect students' strengths and interests,

and value the habits of motivation and effort in their students. All these practices are supported by research into sound practices and are referenced in *Growing Success* (O.M.E., 2010) as being necessary for preparing students for their later careers in a fast-changing knowledge economy. For this reason, it is difficult to imagine the assessment purposes, summative assessment practices, assessment formats or, values and beliefs of the teacher candidates who completed my survey wanting. While the popular media may frequently feature complaints regarding teacher performance, the results from my study suggest that calls for hemlock are hyperbolic.

## References

- Anderson, S.E., & Ben Jafaar, S. (2006). *Policy trends in Ontario education: 1990-2003*. (ICEC Working Paper #1). Retrieved from University of Toronto, Ontario Institute for Studies in Education; <http://fcis.oise.utoronto.ca/~icec/policytrends.pdf>
- Andrade, H. L., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice*, 27(2), 3-13.
- Andrade, H. L., Wang, X., Du, Y., & Akawi, R. L. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *Journal of Educational Research*, 102(4), 287-302.
- Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into Practice*, 48(1), 12-19.
- Ball, S. J. (1998). Big Policies/Small world: An introduction to international perspectives in education policy. *Comparative Education*, 34(2), 119-30.
- Berliner, D. C. (1988, February). *The development of expertise in pedagogy*. Charles W. Hunt Memorial Lecture presented at the Annual Meeting of the American Association of Colleges for Teacher Education, New Orleans. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED298122&site=ehost-live>
- Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, 35, 463-482.

- Berliner, D. C. (2004). Describing the behavior and documenting the accomplishments of expert teachers. *Bulletin of Science, Technology & Society*, 24(3), 200-212.
- Black, P., & Wiliam, D. (1998a). *Inside the Black Box. Raising standards through classroom assessment*. London, UK: King's College London School of Education. Assessment Group of the British Educational Research Association.  
<http://weaeducation.typepad.co.uk/files/blackbox-1.pdf>
- Black, P. J., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-148.
- Black, P., & William, D. (1998c). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.
- Blackmore, J. (2010). Policy, practice and purpose in the field of education: A critical review. *Critical Studies in Education*, 51(1), 101-111.
- Borko, H., & Others, A. (1997). Teachers' developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development. *Teaching and Teacher Education*, 13(3), 259-78.
- Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education*, 7(4), 279-301.
- Brookhart, S. (2007). Expanding views about formative classroom assessment: A review of the literature. In J. McMillan (Ed.). *Formative classroom assessment* (pp. 43-62). New York: Teachers College Press.
- Brookhart, S. M. (2008). Feedback that fits. *Educational Leadership*, 65(4), 54-59.
- Brown, S., McIntyre, D. (1978). Factors influencing teachers' responses to curricular innovations. *British Educational Research Journal*, 4(1), 19-23.

- Brown, S., McIntyre, D. (1982). Influences upon teachers' attitudes to different types of innovation: A study of Scottish integrated science. *Curriculum Inquiry*, 12(1) 35-51.
- Cizek, G. J., Fitzgerald, S. M., & Rachor, R. E. (1996). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment*, 3(2), 159-79.
- Common, D. L. (1983). Teacher power and settings for innovation: A response to brown and McIntyre's "influences upon teachers' attitudes to different types of innovation.". *Curriculum Inquiry*, 13(4), 435-46.
- Darling-Hammond, L. (2000a). How teacher education matters. *Journal Of Teacher Education*, 51(3), 166-173.
- Darling-Hammond, L. (2000b). Reforming teacher preparation and licensing: Debating the evidence. *Teachers College Record*, 102(1), 28-56.
- Darling-Hammond, L. (2005). Educating the new educator: Teacher education and the future of democracy. *New Educator*, 1(1), 1-18.
- Darling-Hammond, L. (2009). Recognizing and enhancing teacher effectiveness. *International Journal Of Educational & Psychological Assessment*, 3, 31-24.
- Darling-Hammond, L., Chung, R., & Frelow, F. (2002). Variation in teacher preparation: How well do different pathways prepare teachers to teach. *Journal of Teacher Education*, 53(4), 286-302.
- Desimone, L. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research*, 72(3), 433-79.
- Doyle, W., & Ponder, G. A. (1977). The practicality ethic in teacher decision-making. *Interchange*, 8(3), 1-12.

- Duncan, C. R., & Noonan, B. (2007). Factors affecting teachers' grading and assessment practices. *Alberta Journal of Educational Research*, 53(1), 1-21.
- Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, 41(10), 1040-8.
- Dweck, C. S. (2010). Even geniuses work hard. *Educational Leadership*, 68(1), 16-20.
- Dye, T. R. (1972). Education: The group struggle. *Understanding public policy* (pp. 131-159). Englewood Cliffs, New Jersey: Prentice Hall.
- Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation*, 28(2), 465-498.
- Frary, R. B., & Others, A. (1993). Testing and grading practices and opinions of secondary teachers of academic subjects: Implications for instruction in measurement. *Educational Measurement: Issues and Practice*, 12(3), 23-30.
- Fuller, F. F. (1969). Concerns of teachers: A developmental conceptualization *American Education Research Journal* 6(2), 2087-226.
- Gaustad, J. (1993). Peer and cross-age tutoring. *ERIC digest*, 79. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED354608&site=ehost-live>
- Ginsburg-Block, M., Rohrbeck, C. A., & Fantuzzo, J. W. (2006). A meta-analytic review of social, self-concept, and behavioral outcomes of peer-assisted learning. *Journal of Educational Psychology*, 98(4), 732-749.
- Ontario Ministry of Education, (2010). *Growing Success: Assessment, Evaluation and Reporting in Ontario Schools: First Edition, covering Grades 1 to 12*. Toronto. Ontario Ministry of Education: Queen's Printer.

- Hattie, J. (1992). Measuring the effects of schooling. *Australian Journal of Education*, 36(1), 5-13.
- Hattie, J. (2009) *Visible Learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY 10016: Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Howlett, M., Ramesh, M., & Perl, A. (2009). *Studying public policy: Policy cycles and policy subsystems*, (3<sup>rd</sup> Ed.). Toronto, Oxford University press.
- Klinger, D. A., Shulha, L. M., Mills, A., & Wade-Woolley, L. (2011). *The foundations of effective assessment in the elementary classroom: A review of the research literature*. Queen's University: Elementary Teachers' Federation of Ontario.
- Kluger, A. N., & Denisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254-284.
- Kremer-Hayon, L. (1991). The stories of expert and novice student teachers' supervisors: Perspectives on professional development. *Teaching and Teacher Education*, 7(5-6), 427-438.
- McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, 20(1), 20-32.
- McMillan, J. H., Hellsten, L. A., & Klinger, D. A. (2011). *Classroom Assessment: Principles and practice for effective standards -based instruction*, (1st Canadian Ed.). Toronto, ON: Pearson Education Canada.

- Nava, F. J. G., & Loyd, B. H. (1992, April). *An investigation of achievement and nonachievement criteria in elementary and secondary school grading*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.
- Pallant, Julie, (2010). *SPSS survival Manual: A step by step guide to data analysis using SPSS 4th edition*. New York, NY 10121-2289, McGraw-Hill.
- Plato, (Jowett, B., Trans.) Posting Date November 3, 2008. Release date February 1999. [Ebook #1656] Retrieved from <http://www.perpustakaan.depkeu.go.id/FOLDEREBOOK/The%20Project%20Gutenberg%20EBook%20of%20Apology%20%28Plato%29.pdf>
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119-44.
- Salkind, N.J. (2008). *Encyclopedia of Educational Psychology*. New York: Sage Publications.
- Scaife, J., & Wellington, J. (2010). Varying perspectives and practices in formative and diagnostic assessment: a case study. *Journal of Education for Teaching*, 36(2), 137-151.
- Schafer, W. D., & Lissitz, R. W. (1987). Measurement training for school personnel: Recommendations and reality. *Journal of Teacher Education*, 38(3), 57-63.
- Scruggs, T. E., & Mastropieri, M. A. (1998). What happens during instruction: Is any metaphor necessary? *Journal of Learning Disabilities*, 31(4), 404-408.
- Sebba, J., Crick, R., Yu, G., Lawson, H., Harlen, W., & Durant, K. (2008). *Systematic review of research evidence of the impact on students in secondary schools of self*

*and peer assessment*. EPPI-Centre, Social Science Research Unit, Institute of Education. University of London. KEEP.

- Sergiovanni, Thomas, J., Kelleher, Paul., McCarthy, Martha, M., Wirt, Frederick, M. (2004). *Educational Governance and Administration* (5<sup>th</sup> ed.). Toronto, On: Pearson.
- Shavelson, R. J., & Stern, P. (1981). Research on teachers pedagogical thoughts, judgments, decisions and behavior. *Review of Educational Research*, 51(4), 455-98.
- Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher* 29(7), 4-14.
- Shepard, L. A. (2010). What the marketplace has brought us: Item-by-item teaching with little instructional insight. *Peabody Journal of Education*, 85, 246-257.
- Stiggins, R. J. (1988). Revitalizing classroom assessment: The highest instructional priority. *Phi Delta Kappan*, 69(5), 363-368.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72(7), 534-539.
- Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice*, 20(3), 5-14
- Stiggins, R. J., & Others, A. (1989). Measuring thinking skills through classroom assessment. *Journal of Educational Measurement*, 26(3), 233-246.
- Tabachnik, G. G., Fidell, L. S., (2007). *Using Multivariate Statistics* (5th Ed.). Boston: Pearson Education.
- Taras, M. (2009). Summative assessment: The missing link for formative assessment. *Journal of further and Higher Education*, 33(1), 57-69.
- Topping, K. J. (2009). Peer assessment. *Theory into Practice*, 48(1), 20-27.

- Topping, K.J., & Ehly, S. (Eds.). (1998). *Peer-assisted learning*. Mahwah, NJ/London, UK: Lawrence Erlbaum.
- Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal*, 27(5), 615-31.
- Tymms, P., Merrell, C., Thurston, A., Andor, J., Topping, K., & Miller, D. (2011). Improving attainment across a whole district: School reform through peer tutoring in a randomized controlled trial. *School Effectiveness and School Improvement*, 22(3), 265-289.
- Umphrey, J. (2009). Toward 21st century supports: An interview with Linda Darling-Hammond. *Principal Leadership*, 10(1), 18-21.
- Utley, C. A., Mortweet, S. L., & Greenwood, C. R. (1997). *Peer-mediated instruction and interventions: Focus on exceptional children*. Denver, CO: Love.
- Vogt, W. Paul (2007). *Quantitative Research methods for professionals*. Toronto, On: Pearson Education.
- Whitebread, D., & Pasternak, D. (2010). Metacognition, self-regulation and meta-knowing. In Littlelon, K., Wood, C., & Staarm, J.K., (Eds.), *International handbook of Psychology in Education*. Bingley: Emerald Group Publishing.
- Wiliam, D., & Leahy, S. (2006, April). *A theoretical foundation for formative assessment*. Paper presented at the American Educational Research Association and the National Council on Measurement in Education, San Francisco, CA.
- Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M., Brandon, P. R., Furtak, E. M., & Young, D. B. (2008). On the impact of formative assessment on student motivation,

achievement, and conceptual change. *Applied Measurement in Education*, 21(4), 335-359.

Young, V. M., & Kim, D. H. (2010). Using assessments for instructional improvement: A literature review. *Education Policy Analysis Archives*, 18(19), 1-40.

## Appendix A: Ethics Approval



March 16, 2012

Mr. Adam Mills, Master's Student  
Faculty of Education, Duncan McArthur Hall  
Queen's University  
511 Union Street  
Kingston, ON K7M 5R7

GREB Ref #: GEDUC-612-12; Romeo # 6006685  
Title: "GEDUC-612-12 Teacher Candidates' Values and Beliefs Surrounding Grading and Assessment Practices"

Dear Mr. Mills:

The General Research Ethics Board (GREB), by means of a delegated board review, has cleared your proposal entitled "GEDUC-612-12 Teacher Candidates' Values and Beliefs Surrounding Grading and Assessment Practices" for ethical compliance with the Tri-Council Guidelines (TCPS) and Queen's ethics policies. In accordance with the Tri-Council Guidelines (article D.1.6) and Senate Terms of Reference (article G), your project has been cleared for one year. At the end of each year, the GREB will ask if your project has been completed and if not, what changes have occurred or will occur in the next year.

You are reminded of your obligation to advise the GREB, with a copy to your unit REB, of any adverse event(s) that occur during this one year period (access this form at [https://eservices.queensu.ca/romeo\\_researcher/](https://eservices.queensu.ca/romeo_researcher/) and click Events - GREB Adverse Event Report). An adverse event includes, but is not limited to, a complaint, a change or unexpected event that alters the level of risk for the researcher or participants or situation that requires a substantial change in approach to a participant(s). You are also advised that all adverse events must be reported to the GREB within 48 hours.

You are also reminded that all changes that might affect human participants must be cleared by the GREB. For example you must report changes to the level of risk, applicant characteristics, and implementation of new procedures. To make an amendment, access the application at [https://eservices.queensu.ca/romeo\\_researcher/](https://eservices.queensu.ca/romeo_researcher/) and click Events - GREB Amendment to Approved Study Form. These changes will automatically be sent to the Ethics Coordinator, Gail Irving, at the Office of Research Services or [irvingg@queensu.ca](mailto:irvingg@queensu.ca) for further review and clearance by the GREB or GREB Chair.

On behalf of the General Research Ethics Board, I wish you continued success in your research.

Yours sincerely,

A handwritten signature in black ink, appearing to read "Joan Stevenson".

Joan Stevenson, Ph.D.  
Professor and Chair  
General Research Ethics Board

cc: Dr. Don Klinger, Faculty Supervisor  
Dr. Lesly Wade-Woolley, Chair, Unit REB  
Erin Wicklam, c/o Graduate Studies & Bureau of Research

## Appendix B: Survey Instrument

### Section 1: Bio-demographic Information

- 1) I am: Concurrent / Consecutive

*Think back to one class you taught during your practicum in an Ontario School.*

- 2) What is the grade level of the class? 7 8 9 10 11 12

- 3) What is the subject of the class?

Math

English

Language other than English

Social Sciences

Science

Physical Education

Art/Music/Drama

Technical Education

- 4) What is the level of the class?

Academic/University

Applied/College

Other

### Section 2: Assessment Purpose

*The following questions are answered on a scale: Not at all, Not Very, Somewhat, Very, Extremely*

- 5) I expect I will assess primarily to:

a) Assign grades

b) Inform my teaching

c) Provide my students with feedback about their progress towards learning objectives

- 6) I expect the majority of my assessments will be used:
- a) for formative purposes
  - b) for summative purposes

### Section 3: Summative Assessment Practice

[Note: Italicized items indicate they were removed during scale reduction.]

*The following questions are answered on a scale: Not at all, Very little, Some, Quite a bit, Extensively, Completely*

7) Thinking about the class you identified above, to what extent do you believe your summative assessments in the future will be based on:

- a) *Including zeros in the determination of final grades*
- b) student attitude toward learning
- c) improvement of performance since the beginning of the semester
- d) low student effort to learn
- e) high student effort to learn
- f) degree of effort of low ability student
- g) overall ability level of student
- h) *major exams or compositions*
- i) oral presentations
- j) work habits and neatness
- k) how much the student tried to learn
- l) whether homework was completed
- m) how well the student did on homework
- n) class participation
- o) *only academic performance*
- p) *improvement by student whose initial performance was very low*
- q) how performance compares to other students that semester

- r) how performance compares to students in previous semesters
- s) how performance compares to set scale of percentage correct
- t) the number and level of specific learning objectives mastered
- u) how much student pays attention in class
- v) how carefully assignments are completed
- w) whether the student seeks help
- x) whether the student seems motivated to learn
- y) assessments that measure student's recall knowledge or measure student understanding
- z) assessments that measure student's reasoning (higher order thinking) or that measure how well they apply what they learn
- aa) *objective assessments (multiple choice, matching)*
- bb) performance assessment (structured teacher observations)
- cc) student response to oral questions during class
- dd) essay type tests
- ee) projects completed by teams
- ff) projects completed by individual student
- gg) performance on quizzes
- hh) using effort, improvement behavior and other "non-test" indicators for borderline cases.
- ii) authentic assessments (tasks close to experiences outside of school/have real world applications)

#### Section 4: Values and Beliefs

8) How important to you is the value of Equity (fair play, equal opportunity, differentiated approaches) in your assessment practice?

Not at all      Not very      Somewhat      Very      Extremely

9) How important to you is the value of Efficiency (reduction in costs of time, money, etc.) in your assessment practice?

Not at all      Not very      Somewhat      Very      Extremely

10) How important to you is the value of Choice (your autonomy as a teacher, your ability to personalize or individualize your assessment practices) in your assessment practice?

Not at all      Not very      Somewhat      Very      Extremely

11) How important to you is the value of Excellence (high standards, the ability of students to have a keen understanding and to perform well) in your assessment practice?

Not at all      Not very      Somewhat      Very      Extremely

12) Thinking of your formative assessment practice please rank the following four values (Equity Efficiency, Choice and Excellence) according to their importance to you. 1-least important, and 4=most important.

- |               |   |   |   |   |
|---------------|---|---|---|---|
| a) Equity     | 1 | 2 | 3 | 4 |
| b) Efficiency | 1 | 2 | 3 | 4 |
| c) Choice     | 1 | 2 | 3 | 4 |
| d) Excellence | 1 | 2 | 3 | 4 |

13) Thinking of your summative assessment practice, please rank the following four values (Equity, Efficiency, Choice and Excellence) according to their importance to you. 1-least important and 4- most important

- |               |   |   |   |   |
|---------------|---|---|---|---|
| a) Equity     | 1 | 2 | 3 | 4 |
| b) Efficiency | 1 | 2 | 3 | 4 |
| c) Choice     | 1 | 2 | 3 | 4 |
| d) Excellence | 1 | 2 | 3 | 4 |

#### Section 5: Assessment Format

14) Please provide an example of a formative assessment tool/task that you feel reflects on your values regarding formative assessment with a brief explanation of why.

15) Please provide an example of a summative assessment tool/task that you feel reflects on your values regarding summative assessment with a brief explanation of why.

Section 6: Background and Experiential Factors Influencing your Assessment

16) How much do you believe the following factors will affect your assessment practices in the future?

a) My beliefs before I entered the B. Ed. program

Not at all      Not very      Somewhat      Very      Extremely

b) My course work in the B. Ed. program

Not at all      Not very      Somewhat      Very      Extremely

d) My teaching practicum

Not at all      Not very      Somewhat      Very      Extremely