UNIVERSITY OF CALGARY

Mathematics Before Teaching: An Analysis of the Academic Transcripts of Prospective Elementary and Secondary Mathematics Teachers

by

Tony Pascuzzo

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF ARTS

GRADUATE DIVISION OF EDUCATIONAL RESEARCH

CALGARY, ALBERTA

JANUARY, 2008

UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "Mathematics Before Teaching: An Analysis of the Academic Transcripts of Prospective Elementary and Secondary Mathematics Teachers" submitted by Tony Pascuzzo in partial fulfilment of the requirements of the degree of Master of Arts.

Supervisor, Dr. Jo Towers
Graduate Division of Educational Research

Dr. Hans Smits
Graduate Division of Educational Research

Dr. Michael Lamoureux
Department of Mathematics and Statistics

24/1/08

Date

Abstract

This study presents an analysis of pre-program academic transcripts of prospective mathematics teachers in a two-year after-degree teacher education program at a western Canadian university. Prospective elementary generalist and secondary math specialist teachers are included. Academic transcripts are analyzed such that findings are presented regarding the topic of and grade received for specific courses. Comparisons are made between admitted and inadmissible subsamples of the secondary math specialist applicants. Findings generally show a great degree of variation in coursework with regard to topics studied. A majority of prospective elementary generalist teachers are shown to have taken no coursework in mathematics. The secondary sample is shown to be composed of distinct groups with regards to coursework. Recommendations are made for altering the admission criteria of the program studied. Observations include that the aggregated measures of coursework and specifications of populations used in previous studies of teacher content knowledge may have serious limitations.

Acknowledgements

I would like to acknowledge the support of family and friends, without which this thesis would not have been possible.

I would like to acknowledge and thank Dr. Hans Smits, the Associate Dean, and the staff of the Division of Teacher Preparation, whose assistance and patience were critical in obtaining and interpreting the raw data.

I would like to thank many faculty and fellow students of education for the countless influential discussions and recommended references which have contributed to my understanding of teaching and learning and my approach to this study.

The support of a number of individual faculty members was critical in the formulation and completion of this thesis. These individuals are thanked for making an effort to exceed their formal and informal supervisory and mentorship roles.

Special thanks are due to my supervisor, Dr. Jo Towers, for her support and tireless editing and feedback. Sincere thanks are also directed to the members of the examination committee for taking the time and energy to review and examine this thesis.

Finally, I would like to acknowledge the teachers that I have had throughout my life, formal and informal, well-liked and not-so-well-liked, as their influence has been instrumental in making me the learner and teacher that I am today.

Table of Contents

List of Tables

CHAPTER ONE: INTRODUCTION

A broad range of theoretical positions hold that the process of learning is centered on the changing of an individual's understanding. Whether transmitted, revealed, constructed, or discovered (or any of a number of other verbs appropriate to the many theories of learning), there is some *thing* that is the object and essence of learning. A teacher's role in such a process is defined in radically varying terms (again, depending on the particular theory held), but it is at least partially determined by the teacher's relationship with the object of learning. That is, as long as one holds that there is a *thing* that is the object of learning, then a teacher's relation to that same object is central to the process of teaching and learning. This thesis begins with the ontological assumption described above.

Understanding how teachers themselves have come to understand their respective disciplines is a grand endeavour, potentially encompassing a broad range of philosophical, theoretical, and empirical questions. All disciplines represented in school curriculum are deserving of the kind of detailed analysis that this study presents. *Mathematics Before Teaching*, as the title implies, confines its scope to the field of mathematics. This merely reflects the author's experience and interest. However, there is no apparent explanation for the undeniable over-representation of the discipline of mathematics amongst studies of teacher content knowledge. It is perhaps the unique nature of mathematics in that it is relatively self-contained as an independent discipline (at least, among the subjects taught in schools: Other subjects make fundamental use of mathematics, whereas mathematics only calls upon other subjects for illustration purposes). Regardless, there is something unique about mathematics that holds the attention of researchers of teacher content knowledge.

Mathematics Before Teaching (a reference to Deborah Ball's influential "mathematics for teaching") is an attempt to better comprehend mathematics teachers' understanding of the prescribed object of teaching: the curriculum. While the variation in curriculum between grade levels and course types is precisely annotated in official documents, the variation in individual teachers' understanding of the curricula is only modestly described in commonly heard statements such as "secondary teachers are subject-area specialists", and "elementary teachers tend to have a relatively limited understanding of mathematics". This study provides a means to quantify and qualify those broad statements such that practical and meaningful observations may be possible.

Measures of teacher knowledge are famously problematic to validate, regardless of whether they are direct measures or proxies. A commonly used proxy for teacher knowledge is the number of relevant post-secondary courses taken by individual teachers. This study relies upon similar course counts, but attempts to bring greater validity by further identifying the topic and grade received for individual courses. It draws upon the application documents of prospective teachers for its raw data, most significantly the academic transcripts of individual prospective teachers. The thesis question may be stated as "What is the nature of the formal post-secondary mathematics education of prospective math teachers, as indicated by frequencies and grades received for different types of mathematics and related courses?" Mathematics courses are included in the study for obvious reasons; related courses are included as they may have provided an opportunity for individual teachers to apply their mathematical understandings, and therefore may have also contributed to the horizontal and vertical integration of their mathematical knowledge.

The thesis adheres to the canonical format by presenting a review of literature, description of the methodology, findings, and discussion and conclusion. The literature review provides a survey of empirical and theoretical efforts to model teacher content knowledge, then empirical studies investigating the effect of teacher content knowledge on student learning, and finally there is a statement of the place of this thesis within the existing literature. The methodology chapter provides a detailed account of how the samples were identified, and data collected and analyzed. The list of findings provides a comprehensive description of all relevant values and interactions for demographic and coursework information. The discussion and conclusion engages in some interpretation of the findings in the institutional context, identification of how the limitations of the study can be used in the interpretation of the findings, and suggestions for future research.

This study will naturally be of interest to teacher educators at the institution where the data was collected, as it provides a comprehensive analysis of applicants' records that has not been attempted previously. Better understanding the coursework history of prospective teachers will allow teacher educators to further tailor the content of the teacher preparation program at the university studied, and potentially to alter the entrance requirements in order to address undesirable coursework trends amongst applicants (should they arise). Teacher educators at other institutions may also be similarly interested in the findings, as the individuals involved in this study may be similar to applicants to and students in their own programs. Math teacher educators in general may be interested in the findings, as detailed studies of the prior coursework of prospective teachers are not prevalent. Finally, professional organizations that provide guidelines for teacher content knowledge may also be interested to see how well prospective teachers' content knowledge aligns with their guidelines.

CHAPTER TWO: LITERATURE REVIEW

Overview

This review of relevant literature is divided into three sections. First, there is a review of theory and research which support the importance of teacher content knowledge as a measure of "teacher quality". This review includes references about the value of teacher content knowledge in general, and content knowledge specifically for math teachers. Second, a collection of research that has focused on the link between teacher content knowledge and student outcomes (primarily student achievement on standardized tests, collectively referred to as "education production function research") is presented. Classic education production function studies are discussed, as well as studies identifying the inherent limitations of this line of inquiry. The third and final section describes the need for research into the nature of the coursework of beginning teachers, and the place of this study within the established literature.

Measures of "Teacher Quality"

The use of the term "teacher quality" here refers to the multiple qualities of a teacher in general, not merely an oversimplified "good" versus "poor" teaching categorization or continuum. Teacher quality, then, encompasses multiple domains of knowledge, skill, and ability. These are not postulated to be discrete and well-defined, but rather overlapping, recursive, and interdependent (e.g. Leinhardt, Putnam, Stein, & Baxter, 1991, p. 87; Grossman, Wilson, & Shulman, 1989, p. 25; Ball, Lubienski, & Mewborn, 2001, p. 453). A full or even partial analysis of the complex relationship between these qualities is beyond the scope of this study, but it is acknowledged that the aforementioned qualities do not operate independently. This study is an attempt to better understand the knowledge of individuals who will be expected to teach mathematics; as such, the unit of analysis is a quantifiable portion of an individual's mathematics knowledge and experience (as illustrated by the nature and number of mathematics and related courses taken by an individual, and the grades received in these courses). The following section details the applicability and the limitations of applying such partial measures.

General measures of teacher quality

Teacher subject-area knowledge appears as a fundamental consideration throughout the history of educational philosophy. It is difficult to avoid, as issues of content and curriculum invariably enter educational debate from all sides. Many approaches to education associate the teacher with some kind of content, some more directly than others. For example, Franklin Bobbitt's or Ralph Tyler's analytic approaches tend to present the teacher's knowledge of content as one of the critical variables to be optimized for maximum student learning (Bobbitt, 1918; Tyler, 1949). Alternately, John Dewey's oft-described "child-centered" philosophy implies a more tangential relationship and presents curriculum (and indeed, teaching itself) more as a means to an end, with teacher knowledge as critical to meeting the changing needs of individual students (Dewey, 1929). More contemporary perspectives on curriculum retain some essence of the fundamental nature of content: William Doll's "Four Rs" of curriculum (R*ichness* - depth of meaning; *Recursion* - complex structures; *Relations* - between curriculum and culture; and *Rigor* - commitment to exploration; presented in Doll, 1993) comprise the criteria by which curriculum should be judged (Flinders & Tornton, 2004, p. 200-201). Satisfying Doll's Four Rs would appear to require a significant grasp of content.

Depending on how one approaches teaching and learning, the importance of content ranges from fundamental (it is the end in itself) to contextual (it is a means to an end, though an *essential* means). At the very least, content forms an inescapable part of the context and process of learning. The link between teacher knowledge of content and effective learning is popular and intuitive: If a teacher "knows" more content, then they ought to be a more effective teacher. This simple claim is surprisingly difficult to verify, and such efforts are hampered by a great number of factors.

Teacher knowledge of the curriculum they are expected to administer is necessarily limited due to a great number of factors. A common observation amongst teachers and teacher educators is that curricula tend to be broad. This is made more problematic in the agglomerative nature of curriculum development: once inserted into curricula, topics tend to remain (Schmidt et al., 1997). Thus, it may not be reasonable to expect teachers to be comprehensive experts of the regularly-changing curricula they are expected to teach. McDiarmid, Ball, and Anderson (1989) state that "Beginning teachers must develop a flexible, thoughtful, and conceptual understanding of their subject matter..." (p. 198). They continue by describing how teachers of various subjects tended to have a shallow view of their disciplines. They state that despite this, teacher knowledge of subject content "rarely figures prominently in teacher preparation," and that "teacher educators tend to take prospective teachers' subject matter knowledge for granted." This is compounded by the fact that elementary generalists tend to believe that deep understanding is not necessary to teach young children, and secondary specialists simply assume that they are experts in high school content since they have studied "beyond" that level. Grossman, Wilson, and Shulman (1989), in the same volume, point out that teacher beliefs about and "orientations" to their discipline are critical in how teachers present the curricula (p. 28-32), and are often formed in educative experiences prior to teacher preparation programs (e.g. in early undergraduate coursework, wherein the substantive structure of a discipline may or may not be explored). McDiarmid, Ball, and Anderson (1989) continue by noting that subject mastery is "beyond the reach of any single course" (p. 199) but is an issue that teacher educators cannot ignore.

Official sources echo this sentiment. The USA's National Council for Accreditation of Teacher Education (NCATE) states "Candidate Knowledge, Skills, and Dispositions" as their first *Standard*: "Candidates preparing to work in schools as teachers ... know and demonstrate the content, pedagogical, and professional knowledge, skills, and dispositions necessary to help all students learn" (National Council for Accreditation of Teacher Education, n.d.). In implementing nationwide teaching standards in the USA, the National Board for Professional Teaching Standards (NBPTS) similarly state the centrality of teacher content knowledge in their *Proposition 2*: teachers "have mastery over the subject(s) they teach. They have a deep understanding of the history, structure, and real-world applications of the subject" (National Board for Professional Teaching Standards, n.d.). The Association of Canadian Deans of Education (ACDE) makes a similar claim in supporting as a *Principle of Initial Teacher Education* that "An effective teacher education program ensures that beginning teachers have sound knowledge of subject matter, literacies, ways of knowing, and pedagogical expertise" (Association of Canadian Deans of Education, n.d.). Subject-specific teacher organizations go further, by providing specifying content expectations in professional standards for teaching within the subject area (e.g. National Council of Teachers of Mathematics, 1991). In the local context, the "Descriptors of Knowledge, Skills and Attributes Related to Interim Certification" (KSAs) composed by Alberta Education specifies that "teachers ... are expected to demonstrate consistently that they understand... the subject disciplines they teach. They have completed a structured program of studies through which they acquired the knowledge, concepts, methodologies and assumptions in one or more areas of specialization or subject disciplines taught in Alberta schools" (Alberta Education, n.d.).

Teacher knowledge is thought to have critical importance for teaching practice. An influential exploration of the nature of such knowledge is presented in Lee Shulman's 1986 article, "Those who understand: Knowledge growth in teaching". After asserting that subject content is a "missing paradigm" in contemporary teacher education programs (Shulman, 1986, p. 6-10), Shulman describes three categories of teacher knowledge: "content knowledge", or context-free generalist knowledge of the subject that an expert in the area would have; "pedagogical content knowledge", or specialized knowledge of the subject that is specific to teaching and learning; and "curricular knowledge", both vertically-oriented knowledge of the curriculum being taught and a lateral knowledge of other curricula concurrently studied by the students (for the purpose of integrating learning across subjects). Shulman states that these three domains of teacher knowledge are represented via three "forms" for representing that knowledge: Propositional, case, and strategic knowledge (Shulman, 1986, p. 10 – 13). This study makes use of measures that belong to Shulman's conceptualization of the content knowledge domain: The fundamental importance and interdependency of all three categories of knowledge is recognized, but the measures used are firmly bounded within the "content knowledge" domain.

Shulman defines content knowledge as "the amount and organization of knowledge per se in the mind of the teacher" (Shulman, 1986, p. 9). He refers to Bloom's cognitive taxonomy in asserting that content knowledge ranges from simple facts to complex applications. He also invokes Schwab's distinction between substantive and syntactic structures, stating that both types are incorporated into teacher content knowledge (Schwab, 1978). In summary, Shulman states that the teacher should possess both a comprehensive and deep understanding of the content, so much that "the subject matter content understanding of the teacher be at least equal to that of his or her lay colleague, the mere subject matter major" (Shulman, 1986, p. 9). Theories of content knowledge specific to math teachers are detailed in the next section.

The critical importance of subject matter knowledge is well illustrated by Grossman, Wilson, and Shulman (1989, p. 28) in two parallel examples. They recount an instance in which an English teacher was informed of her teaching assignment on the very day that classes began. She was to teach *Beowulf*, a book that she had not read previously. Lacking time to prepare, she was forced to simply keep a few pages ahead of her students; in doing so, she noted that she was unable to address student questions about the plot, foreshadow important events, and emphasize recurrent themes. Similarly, they state that a math teacher without a deep understanding of geometry might not be able to identify or emphasize key concepts as they arise, or draw attention to and illustrate fundamental relationships between problems and concepts.

It is difficult to precisely quantify the amount and degree of experience that an individual learner has with a particular topic. In an attempt to help understand the relationship between experience with and corresponding understanding of a topic, "opportunity to learn" (OTL: the amount of time spent on a particular topic in the classroom) has been established as an important measure of student learning. Schmidt et al. found that OTL was a significant predictor of student achievement in the TIMSS 1995 study (Schmidt et al., 2001, p. 339-346). It is likely that OTL is equally important for teacher learning, as prospective teachers have a finite amount of time available in which to learn about their subject. If a teacher were not to have sufficient opportunity to learn about a particular topic, it may be the case that they will not be able to teach it as effectively as if they did. Increasing OTL was found to have the greatest effect on student achievement in subject areas where students have relatively little educational exposure.

This is consistent with the intuitive notion that one can make the greatest overall gains by improving upon one's weaknesses. In the context of prospective teachers, this implies that individuals ought to focus on areas in which they have little or no exposure. It is unknown if the model suggested above (in observing OTL for students) may be appropriately extrapolated to the case of prospective teachers (as it may be argued that the goal of teacher preparation is not to increase overall teacher achievement, or anything similar), but the similarities and implications are not insignificant. In either case, detailed information about the particular topics to which prospective teachers have had exposure in their previous degrees will be of interest to teacher educators.

Math-specific measures of teacher quality

A model of teacher knowledge similar to Shulman's widely recognized conception (Shulman, 1986) was postulated specifically for teachers of mathematics by Deborah Ball (1990). Her landmark article provided an argument and evidence for the existence of "mathematics knowledge for teaching" (which has profoundly affected the field of mathematics teacher education – its influence also extends to the title of this thesis), and challenged three common assumptions about learning to teach mathematics: "(1) that traditional school mathematics content is not difficult, (2) that precollege education provides teachers with much of what they need to know about mathematics, and (3) that majoring in mathematics ensures subject matter knowledge" (Ball, 1990, p. 449). Ball's model of "knowledge for teaching" was directly compared to Shulman's knowledge domains, with general alignment (Ball et al., 2005). Ball's model was asserted to be a further elaboration of Shulman's conception within two of Shulman's original domains: "Pedagogical content knowledge" (Shulman, 1986) was divided into "knowledge of content and students" (KCS) and "knowledge of content and teaching" (KCT), and "content knowledge" (Shulman, 1986) was divided into "common content knowledge" (CCK) and "specialized content knowledge" (SCK). CCK was defined as the mathematical knowledge that any person able to perform a given task requires (e.g. applying correct procedures and notation, recognizing incorrect statements, etc.), whereas SCK was defined as the additional knowledge required in order to effectively teach the same task (e.g. identifying and explaining concepts, analyzing and giving feedback on incorrect statements, using alternate representations, etc.). Both CCK and SCK are critical to sound teaching practice, though only CCK is necessary for success in typical undergraduate mathematics programs.

Davis and Simmt (2006), in direct reference to Ball's model, note that "mathematics knowledge for teaching is likely neither a matter of "more of" nor "to a greater depth than" the knowledge expected of students. It is qualitatively different" (p. 294). This claim of the existence of a qualitatively different type of mathematics knowledge is echoed in Liping Ma's specification of "profound understanding of fundamental mathematics": a vertically and horizontally integrated knowledge of conceptual mathematics (Ma, 1999). The unique nature of the mathematics knowledge necessary for teaching is postulated to comprise an entire sub-discipline within mathematics (Usiskin, 2001). It is claimed that "teachers' mathematics" is a distinct branch of applied mathematics (Usiskin, 2001, p. 14):

> Like other branches of applied mathematics, teachers' mathematics uses only a part of all known mathematics, and favors certain areas. Number theory, geometry, and the foundations of mathematics are important to teachers' mathematics in the same way that probability is important to actuarial science... Also, like other branches of applied mathematics, knowledge in non-mathematical areas is very helpful. Thus, just as a financial analyst needs to know about the various investment possibilities available... a teacher needs to know about learning theory, student motivation, and the effects of schooling and testing on student learning.

In a perhaps surprising parallel, influential curriculum theorist and evaluation pioneer Ralph Tyler commented on the differences between the knowledge of mathematics required by mathematicians, and that which is useful in public schools: "What can [a] subject contribute to the education of young people who are not going to be specialists in your field?" (Tyler, 1949, p. 56). While not directly suggesting that math *teacher* knowledge should be different from that of the mathematician, he is quite clear that the content being taught in schools should be substantially different from what would be appropriate for future mathematicians.

Empirical support for the distinct nature of teacher knowledge in mathematics is presented in a 2004 article by Hill, Schilling, and Ball. They note that there is little agreement over the substance of knowledge of mathematics needed for teaching (Hill, Schilling, & Ball, 2004, p. 12). The authors state that the purpose of their study was twofold: to determine whether the "mathematics knowledge for teaching" was a single or multi-dimensional construct, and whether or not such knowledge could be reliably measured. The researchers developed a test instrument designed to illustrate the differences between general mathematical knowledge (which individuals without a significant pedagogical background would be expected to know), and mathematical knowledge for teaching (based on theoretical constructs identified in the literature). Their results provided support for the claim that general and pedagogical content knowledge are related but not entirely equivalent. It also was fairly conclusive in supporting the claim that teacher knowledge of mathematics for teaching is at least partly domain specific (rather than being purely related to overall intelligence, mathematical ability, or pedagogical skill). Overall, the study provides empirical support for the theoretical models of teacher knowledge that differentiate mathematical knowledge for teaching from general mathematical knowledge: for example, Shulman's "pedagogical content knowledge" (Shulman, 1986), and Ball's "mathematics for teaching" (Ball, 1990).

Summary

Many researchers have established the relevance of teacher subject-area knowledge as one of the many measures of "teacher quality". Ball, Lubienski, and Mewborn (2001) make an illustrative observation: "The claim that teachers' knowledge matters is commonsense. However, the empirical support for this 'obvious' fact has been surprisingly elusive... No consensus exists on the mathematical knowledge that is required to teach" (p. 441). Indeed, there is little consensus beyond the general notions that *some* mathematical knowledge is helpful for teachers, and that more is probably better but only to a certain threshold.

Various organizations (each of them predominately American) have provided detailed guidelines as to the specific content that math teachers should be comfortable with (or guidelines for student content for which it may be assumed that teachers should also be comfortable with). These include, but are not limited to, the National Council of Teachers of Mathematics, the American Mathematics Society, the National Assessment of Educational Progress, the Third International Mathematics and Science Study, and Educational Testing Systems. In an attempt to answer the question of "What mathematics should elementary grade mathematics teachers really know?" McGatha, Brown, and Thompson (2007) reviewed eleven such student and teacher guidelines available. They found considerable areas of agreement amongst the topics included in the separate guidelines, but were troubled by the highly variable emphasis given to different topics across the eleven guidelines (McGatha, Brown, & Thompson, 2007). As their research was confined to the topics that elementary teachers ought to know, it may be the case that greater variability and less overall alignment would be found for the same analyses of secondary math topics (as the curricula generally become more detailed and in-depth at the high school level).

In reference to the origin of the types of standards documents discussed immediately above, Ball, Lubienski, and Mewborn (2001) comment: "...such lists usually identify topics beyond the curriculum. This approach to specifying knowledge for teaching is rooted primarily in policy deliberations and often does not reflect research evidence" (p. 441). It is accepted that individual knowledge is necessarily finite: "Even at the secondary level we anticipate that prospective teachers would not have studied all of the mathematical domains that might be required in their teaching" (Sullivan, 2003). The question remains of how much content knowledge is helpful, and which particular topics? Thus, it is hoped that research into the types of mathematics that prospective teachers *do know* will be of service in determining the types of mathematics that prospective teachers *ought to know*.

Education Production Function Research

As stated in the previous section, there is little empirical evidence for the direction and degree of the relationship between what Ahn and Choi (2004) describe as "teacher quality and quality teaching" (that is, the measurable characteristics of teachers – teacher background variables, and the measurable outcomes of teacher practice – student achievement scores), regardless of how one chooses to specifically define these terms. Many studies have proposed using prior subject-related coursework (defined as number and nature of courses taken, and grade received) as one of the many possible measures of teacher quality, and this study does the same. The relationship between teacher variables and student achievement has often been investigated using the "education production function" paradigm (Hill, Rowan, & Ball, 2005, p. 373). That is, some collection of characteristics or behaviours of teachers are related to student outcomes (namely achievement) with the intent of identifying positive and negative teacher variables that could be manipulated in order to improve student achievement. The intent is to demonstrate the effect of such variables, and to improve the state of education by informing empirically-driven policy changes.

What would seem to be a simple process of controlled experimentation (or, if not simple, then well-understood) has historically been confounded by the complex nature of teaching and learning, resistant to being "pinned down" as a collection of first-order cause-and-effect relationships. The following studies are representative of those in the area, and retain their influential positions as regularly-referenced literature.

Classic studies

The classic approach to the education production function is experimental: all other things being equal, does a teacher's greater knowledge of content result in greater student learning? In 1986, a survey of 147 existing education production functions (although these all arose from 38 distinct studies - most provided multiple production functions) was conducted to establish the economic efficiency of school expenditures (Hanushek, 1986). Variables that were known to contribute to school expenditures were correlated to student achievement (on a variety of standardized tests). Hanushek (1986) states, "The results are startlingly consistent in finding no strong evidence that teacher-student ratios, teacher education, or teacher experience have an expected positive effect on student achievement" (p. 1162). The author notes that while six studies found a significant and positive relationship between teacher education and student achievement, five studies found a significant negative relationship. The vast majority found no statistically significant relationship (95 studies). Hanushek states that one should be cautious in interpreting the findings, as they rely on the veracity of the original studies. Additionally, he illustrates how contextual factors not accounted for in the original studies could easily mask the identification of any relationship (e.g. not controlling for student ability or the tendency for remedial classes to be smaller in size; Hanushek, 1986, p. 1193).

Hanushek's study is regularly quoted in academic literature on the subject. It represents one of the first attempts to analyze production function studies collectively. However, it was demonstrated to be methodologically inferior by Hedges, Laine, and Greenwald, in an exchange on the pages of *Educational Researcher* (described in Crawford & Impara, 2001, p. 137-138, and Brewer & Goldhaber, 1996, p. 248-249). For the purposes of this literature review, it is sufficient to note that Hanushek's method gave undue weight to poorly designed original studies, applied an over-simplified method, and that a meta-analysis of the same original studies found a different (positive and significant) relationship (Crawford & Impara, 2001; Brewer & Goldhaber, 1996). While Hanushek's study has its share of flaws, it retains significant influence, and is described here as an illustration of the perhaps surprising wealth of inconclusive studies on the subject.

More recent empirical work using more sophisticated data and modeling has been relatively successful. In a quasi-experimental study of 5,149 tenth-grade students in American public schools (conducted using data from the National Educational Longitudinal Study of 1988), Brewer and Goldhaber were able to construct a relatively sophisticated model of educational productivity that addressed dozens of individual, family, school, class, and teacher variables. One of the most significant developments in this study was the use of longitudinal student achievement data: student *improvement* on the tenth-grade mathematics test was used as the output variable. Previous studies typically were not able to control for student ability. Brewer and Goldhaber found that a teacher's post-secondary degree (Bachelor or Master) had a significantly positive effect on student math achievement gain scores only if the degree was in mathematics. When the major area of the teacher's degree is not differentiated, the effect is insignificant. In fact, teachers holding degrees in other disciplines were shown to have a significant negative effect on student math achievement gain scores (Brewer & Goldhaber, 1996, p. 261).

Brewer and Goldhaber's study is instructive for several reasons. First, it provides robust empirical support for the intuitive notion that teacher background and content knowledge has a positive effect on student learning. Second, it permits some explanation of why less sophisticated models fail to demonstrate a significant relationship between teacher variables and student learning (e.g. not differentiating between major area of teachers' degrees obscures the fact that subject-related degrees are a benefit, at least in mathematics). Finally, it demonstrates a necessary condition for appropriate education production function research: Significant attention must be given to developing a sufficiently sophisticated model in order to discern the subtle and complex relationships inherent in teaching and learning.

The final representative study discussed provides additional empirical support for the positive effect of teacher content knowledge on student achievement in science and mathematics using the Longitudinal Survey of American Youth (LSAY) dataset. Mindful of the limited successes of previous studies in the area, Monk (1994) states that "more refined analyses are capable of demonstrating substantive effects of an administratively manipulable schooling attribute – namely, the subject matter knowledge of teachers" (p. 126). Similar to Brewer and Goldhaber, Monk attempted to identify a relatively sophisticated model by addressing dozens of teacher and student variables.For teachers having five or fewer math courses, the effect of having taken additional courses was associated with a 0.2% increase in student math test scores at the grade ten level, and a 1.2% increase at the grade eleven level. Beyond five math courses, the effect was seen to diminish such that Monk described the relationship as curvilinear or having a threshold property. In a tangential inquiry, he found that the benefit of teachers' additional coursework was pronounced for students in advanced courses (enrichment-type courses), and nil for students in remedial courses (further analyses showed that this differential effect arose not because of "student ability", but rather because of the nature of the content being taught in the classes). In a discussion of the findings, Monk concluded that "a good grasp of one's subject areas is a necessary but not a sufficient condition for effective teaching" (p. 142).

While Monk was successful in applying "more refined analyses" (p. 126), the study used several measures that might have benefited from disaggregation. In reporting the number of courses taken in science, teachers were asked to distinguish between biology, chemistry, physics, and earth science courses; In mathematics, no such distinctions were made, though the theoretical applicability of a university course in statistics to teaching high school geometry is certainly no stronger than that of a university course in biology to teaching high school physics. Further distinctions within the field of mathematics might have revealed more robust findings. Similarly, student achievement was determined using composite test scores for math and science separately. Linking student scores on particular topics with greater resolution (e.g. geometry, functions, etc.) to the previous coursework of teachers similarly defined (e.g. calculus, linear algebra) might have been tremendously insightful.

Criticism and meta-analysis

The education production function paradigm arose from taking an economic perspective on the process of education (Brewer & Goldhaber, 1996). The limitations of this approach are detailed by many researchers in the field of education. In a description of existing process-product research (synonymous with education production function research), Shulman states that "In their necessary simplification of the complexities of classroom teaching, investigators ignored one central aspect of classroom life: the subject matter" (1986, p. 6). The net effect was that the findings of such studies presented the process of teaching as a generalized phenomenon, where issues and effects peculiar to individual subjects could not be identified. Shulman's observation of the simplification (and sometimes over-simplification) inherent in this type of research encapsulates the bulk of the criticism of the body of research. While it is arguable that any attempt to describe or model the process of teaching and learning entails some amount of simplification (simply on the basis that one could not hope to address all possible contextual variables with sufficient detail to capture their entirety), even the most sophisticated models in education production function research still primarily rely on aggregate proxy measures of undetermined or poorly defined validity.

In general, existing studies of the effect of teacher quality on student achievement are noted for their inadequacies (Darling-Hammond, 2000; Rowan, Correnti, & Miller, 2002; Hill, Rowan, & Ball, 2005). Criticisms are directed at measures of student learning (usually scores on standardized tests), measures of teacher quality, and period of teacher effect. The validity of these measures is critical to the education production function debate.

Darling-Hammond (2000) notes that "teacher effects [on student achievement] appear to be additive and cumulative, and generally not compensatory" (p 2). Naturally, this presents a significant problem for studies into teacher effects, as the measures are difficult to administer even once, let alone the several times necessary to collect the longitudinal data referenced by Darling-Hammond. Rowan, Correnti, and Miller (2002) also note that typical studies of teacher effectiveness use the student achievement outcome measured at one particular point in time or over a relatively short span, thus obscuring potential cumulative effects (p 1528). They note further issues with typical studies using student achievement as measures of teacher effectiveness: while student achievement represents a cumulative construct (learning having occurred over many years), the effect of a particular teacher on students occurs only over the course of one year or less (Rowan, Correnti, & Miller, 2002, p. 1541).

Hill, Rowan, and Ball describe many shortcomings of the education production function literature. Foremost of these is the reliance on proxy measures for teacher knowledge, such as teacher tests and qualifications (Hill, Rowan, & Ball, 2005, pp 375-376). The authors state that the decontextualized and simplified measures which are often used are not adequate measures of teacher knowledge. This sentiment is repeated in Wilson, Floden, and Ferrini-Mundy's (2002) review of 313 studies in teacher preparation research: "We found no reports meeting our selection criteria that directly assessed prospective teachers' subject matter knowledge and evaluated the relationship between teacher subject matter preparation and student learning" (p. 191). They continue by expressing a lack of faith in the validity of the proxies that are used as measures of subject matter knowledge, noting the contradictory or inconclusive findings that are the product of research relying on such proxy measures.

In a vice-presidential address to the measurement and research methodology division of the American Educational Research Association, Reckase (2007) noted that student achievement as measured on tests is a gross oversimplification of student learning, which is itself an indirect measure of teacher quality. He continued in stating that "test scales represent hypothetical constructs", and as such are subject to the common psychometric issues of representative content sampling, and differential scale functioning for various known and unknown reasons (e.g. low scoring vs. high scoring, multidimensionality, test-taking behaviour like guessing and non-response strategies; Reckase, 2007). Each of these reasons alone would be sufficient to reject analyses based on the composite and aggregate measures typically used in education production function research. Reckase concluded by stating that these issues in simultaneous interaction could result in an unquantifiable compound error of measurement. This provides a strong argument against the use of aggregate and composite measures of teacher quality and student learning: A greater level of detail and resolution (e.g. looking at how teacher knowledge affects student achievement within particular topics in math: geometry, or functions, for example) is thus appropriate for investigating this relationship.

The methods of meta-analysis are suggested in trying to synthesize the extensive, varied, and apparently contradictory education production function studies. Crawford and Imapara (2001) note the importance of being able to appropriately compare and combine "methodologically excellent studies" with "less-than-excellent studies" to determine their relative importance and contribution to the field (p. 159).

Ahn and Choi conducted an extensive synthesis of 41 studies measuring the relationship between teacher quality and quality teaching (Ahn & Choi, 2004). Here, teacher quality was defined as "subject matter knowledge" as measured by both teacher educational background variables (e.g. coursework, GPA, certifications) and performance on teacher tests. Quality teaching was defined as how well students of these same teachers performed on standardized tests. They observed great variance in the results of such studies, noting that the results range from strong positive to moderate negative associations. Ahn and Choi undertake a meta-analysis of a subset of 27 of these studies for which reasonably comparable measures of association were available (i.e. those in which a correlation coefficient was either given or was able to be calculated). They found that the overall association between teachers' subject matter knowledge and student achievement was positive, but very small ($r=.06$). This association was slightly lower for teachers and students at the elementary level and slightly higher at the secondary level. Furthermore, they found that teachers' subject matter knowledge specifically as measured by coursework variables had a stronger association with improved student performance at the secondary level. Finally, it should be noted that the authors performed a test to detect publication bias amongst the studies included in the meta-analysis. A funnel plot revealed a satisfactorily symmetrical distribution of findings amongst the included studies, indicating no obvious publication bias. This provided additional confidence in their findings, as negative publication bias (the systematic non-reporting of inconclusive or negative results) can significantly skew the results of meta-analyses.

There were several limitations of Ahn and Choi's study noted in the paper itself. First, only the studies wherein a correlation coefficient was reported or could be calculated were included in the meta-analysis. Unfortunately, this meant that many studies were excluded on the basis of methodology, such as those using analysis of covariance, hierarchical linear modeling, and multiple regression. Qualitative studies of the same association between teacher quality and quality teaching were likewise excluded. Also, the authors use student achievement test scores as the sole measure of quality teaching, which they acknowledge as only part of the larger picture of student learning. The authors included all possible correlation observations from the studies, and thus over-representation of certain populations and methodologies were possible (i.e. as many as nine correlation coefficients from a single study are included, and given equal weight to other individual correlation coefficients). Finally, the authors note that their selection of variables used to identify moderating effects is not exhaustive, and that others may hold great explanatory power. For example, it may be instructive to test the difference in effect for teacher quality on student achievement on large-scale standardized tests versus local or teacher-constructed tests (as the latter may be more closely aligned with what the teachers taught). In another example, the authors suggest that the effect of teacher quality may vary according to the particular content area tested, or due to the relative academic ability of the students tested.

Ahn and Choi's study provides general support (albeit weak support) for the assertion that teacher quality as measured by academic background is positively associated with student achievement. In reviewing their findings, it is reasonable to assert that the inconsistency noted in the literature may be due to the fact that individual studies use different measures of teacher quality (amount and quality of coursework vs. teacher tests), at different levels of schooling (elementary vs. secondary), with student populations of different demographic and achievement variables (random national sample vs. single school), and different content sampling (comprehensive math content sample vs. one single math topic). Ahn and Choi's inclusion of all such studies with equal weight given to each reflects the nature of such meta-analyses. They note that none of the moderating variables identified completely explain the variance in the findings of the included studies, nor do combinations of said variables. This implies that variables other than those identified are responsible for at least part of the variance in the findings.

The place of education production function research

There is a fundamental problem in trying to pin down education production functions: Clear and interpretable results will likely only be obtained in truly experimental studies involving pretest, treatment, and post-test, all with a control group. Existing studies tend to rely on quasi-experimental design (i.e. not pre- and post-test format, necessarily testing multiple variables simultaneously), prone to confounded results by a proliferation of complex, poorly understood, and inherently difficult to quantify factors which are proxy measures of unknown validity. The question of teacher knowledge might better be approached via simple, single-order relationship. Does a teacher's "greater" understanding of a particular math topic or concept (deeper, more horizontally and vertically integrated, more sophisticated) result in similarly "greater" student learning of that same topic? This topic-level impact might be subtle such that it is not observable using typical composite measures beyond the classroom level (being confounded and obscured by countless other cumulative and aggregate effects).

Regardless of the limitations of the education production function paradigm, it may demonstrate practical relevance given sufficiently sophisticated models (e.g. Monk, 1994; Brewer & Goldhaber, 1996). Crawford and Impara (2001) note these same inadequacies, yet maintain that "we must pay attention to the education-production-function debate" (p. 159-160) because of its central importance in dominant policymaking discussions. Similarly, Floden (2001) notes that education production function studies will likely become more sensitive to the subtleties of the teaching and learning process: "Although the complexity of the education system often makes such effects difficult to identify, the demands for evidence will encourage investigators to contrive research designs that will deliver" (p. 14).

The Place of *Mathematics Before Teaching*

The topics that form the content of post-secondary mathematics education are not particularly well aligned with the mathematics that teachers will be expected to teach at both the elementary and secondary level (Grossman, Wilson, & Shulman, 1989; Luk, 2005; Usiskin, 2001). Both Shulman (1986, p. 8) and Usiskin (2001, p. 2) observe that there is great potential for secondary teachers to be expected to teach a topic that they have not had any exposure to since their own high school education. It should be noted that it is probably not reasonable for individual teachers to be expected gain the depth of mastery of *all* relevant content called for in the literature (given the broad scope of typical curricula at both the elementary and secondary levels). Field (2006) states that in contrast to the dominant "cult of presentism", it is not a reasonable expectation for teacher preparation programs to produce "finished" teachers (i.e., those whose knowledge of their disciplines is complete). Indeed, it has not been established that master teachers with extensive experience possess a "finished" knowledge of content. If this were the case, the definition of the content knowledge that teachers *should have* would be much more easily attained (given a model of "finished" content knowledge).

Curricular content remains inescapable in teacher preparation. As noted previously, at the very least (or most profound) it forms a part of the essential context of learning. While one might agree that teachers cannot be expected to master all relevant content, there is an expectation that teachers should be familiar with *some* content. Complementary to the common question of "What math teachers *should* know" is the question of "What math teachers *do* know": Teacher educators should be aware of the necessity of establishing the existing knowledge and understandings that students bring to the learning experience. This study, *Mathematics Before Teaching*, presents an investigation of one of the possible measures of teacher content knowledge: the number and type of mathematics and related courses taken by prospective math teachers.

There was a surprising lack of research of this type noted in conducting this review of the literature. No studies were found that attempted to explore the nature of mathematics courses taken by teachers and prospective teachers beyond the distinction between undergraduate and graduate level courses. In a survey of studies on the impact of subject-specific study, Floden and Meniketti (2005) make a similar observation: "No studies address questions about different combinations of college mathematics courses" (p. 283). This is echoed by Zeichner (2005, p. 749) in defining a research agenda for teacher education:

> There are whole aspects of teacher education that remain virtually unexplored by researchers and need careful study. These include the nature and impact of subject matter and general education preparation of teachers… Given that much of the coursework of prospective teachers is taken outside of education schools and departments, it is important for researchers to broaden the scope of research to include these neglected areas.

In Floden and Meniketti's (2005) survey, the authors make three recommendations for strengthening the empirical basis for studying the effect of teacher knowledge of subject matter: (a) improving measures of teachers' knowledge; (b) making use of national and international data sets; (c) sharpening the vocabulary for describing college coursework (p. 284-287). This thesis contributes to all three of these recommendations in different ways. First, this thesis will present a more detailed understanding of the types of courses that prospective teachers have taken, which will help to improve measures of teachers' knowledge. Incidentally, Zeichner (2005) makes a similar recommendation "for strengthening research in teacher education… Development of better measures of teacher knowledge and performance" (p. 740). The currently available studies of course-level information are few in number and are restricted to studies of individual courses, and are relatively context-bound (Floden & Meniketti, 2005, p. 283). Applicability to the second recommendation is more oblique: Floden and Meniketti detail that in regards to the investigation of large-scale data sets, "work at the course level will likely be more productive" than attempts to attribute teachers' learning to entire programs (p. 286). This thesis will begin the process of identifying useful course-level variables, such as the particular topics studied. Finally, this thesis will help to "sharpen the vocabulary for describing college coursework" (p. 286) in mathematics, as Floden and Meniketti note that the established distinctions are too broad for more sophisticated analyses. This thesis will apply a more detailed approach to the classification of the coursework of prospective teachers, identifying categories of math and related courses as they emerge from the sample. Similar findings of a lack of research in the area and recommendations for future studies are found in Wilson, Floden, and Ferrini-Mundy's 2001 report for the U.S. Department of Education entitled "Teacher Preparation Research: Current Knowledge, Gaps, and Recommendations".

To conclude this chapter, this study assumes the widely accepted notion that some subject knowledge is necessary for teacher effectiveness. It is an investigation of the variation in subject knowledge that prospective mathematics teachers have; it is descriptive, and may serve to better illustrate the nature of the subject knowledge held by applicants to the teacher preparation program studied. It is hoped that research into the types of mathematics that prospective teachers *do know* will be of service in determining the types of mathematics that prospective teachers *ought to know*.

CHAPTER THREE: METHODOLOGY

Data was collected by referencing the application records of individuals who applied for entry to the Master of Teaching Program at the University of Calgary. It is intended to provide a comprehensive illustration of the mathematics-related academic history of applicants to the program, while acknowledging and working within the constraints of the source data. As such, the study is most accurately described as exploratory, empirical, descriptive, and post hoc (Cohen et al, 2005).

Populations and Samples

The first step was to identify the populations of interest, and to devise an adequate sampling regime. As the study intends to describe the math-related academic history of math teachers, all individuals who applied to the MT Program who would also be expected to teach mathematics as a regular part of their teaching assignment were included in the population. Within the program, there are four distinct streams that prepare prospective teachers who will be expected to teach mathematics: secondary school mathematics specialists, elementary generalists, early childhood education specialists, and French immersion specialists. These groups are independent, as applicants may belong to only one of the groups at any given time.

Secondary school mathematics specialist sample

Secondary school math specialists were included, as they are commonly recognized as being "subject matter experts", and are thus of central interest to a study of math teacher subject knowledge. The number of applicants enrolled in the secondary mathematics program varies from year to year, from a minimum of 11 to a maximum of 22 over the eight years that the program has been in operation. A total of 116 applicants were enrolled in the secondary mathematics program in that period. This was identified as the population of secondary mathematics specialists. According to Krejcie and Morgan (1970, p. 607-608), a sample size of 92 would be representative of such a population. Cohen et al. state that one should use relatively large samples when there is great heterogeneity within the population over the variables studied (2005, p. 94). Given the tremendous variation in academic history that is possible with these individuals (from minimally meeting the admission criteria through to advanced graduate work in the area), all available records were included. Additionally, secondary math specialist applicants who were unsuccessful or those who declined offers of admission were identified as a distinct population. Unfortunately, the records of these individuals are only briefly kept on file. All 60 of the available records were included (for the same reason as above), including all such applicants over the last two years of the program's operation.

Elementary school generalist sample

Elementary generalists were included, so as to provide a representative overview of the math teaching that occurs in elementary schools. This portion of the MT Program admits the largest number of students: almost 200 individuals on a yearly basis. An adequate sample of all applicants over the eight years of program operation would require rather complicated stratified systematic random sampling (Borg & Gall, 1989, p. 224-225). Additionally, the author observed that investigating the academic history of more recent applicants could be more instructive, so the decision was made to focus on recent years of program operation. According to Krejcie and Morgan (1970, p. 608), a sample size of 196 would suffice for a population of 400, 234 for 600, and 260 for 800. In exercising prudence regarding the representativeness of the sample (Cohen et al, 2005 p. 93), it was determined that the inclusion of only two years of applicants, giving a population of 400, would likely not be sufficient to describe elementary math teaching. Three years was determined much more likely to be sufficient, giving a population of 600 with a sample of 234. This was chosen over the inclusion of four years of records because it would allow the use of a purely systematic random sampling method (Borg & Gall, 1989, p. 224). The sampling method used was simply to include one out of every two records; this had the great advantages of being clear, easy to execute, and not prone to error (e.g. double counting, or systematic error). It also had the advantage of providing a sample size well in excess of that specified by Krejcie and Morgan (1970, p. 608), which was again desirable as the heterogeneity within the variables of the study were expected to be quite high. Thus, one can be more confident in making observations within this sample.

Not included in the study

While they would certainly be interesting groups to study, the early childhood education and French immersion specialist groups were not included due to logistical constraints. The French immersion group represents approximately 4% of student teachers within the identified population (those expected to teach math), so their inclusion was not deemed reasonable given the considerable investment of time in adequately sampling and including their records.

Similarly, the early childhood education specialists were not included; while they comprise a significant portion of the identified population (approximately 16% on a yearly basis), they are representative of mathematics teaching at the K-3 level only. Adequate sampling of their records would have increased the total number of records required at the elementary level by about 32%, based on the inclusion of 92 additional records, the portion (specified in Krejcie & Morgan, 1970, p. 608) of the approximately 120 records. The elementary generalist population comprises the majority of those responsible for teaching math at the elementary level. The relevance to the early childhood education specialist population cannot be established, as this group may have unique characteristics that significantly differentiate it from the elementary generalist sample.

Generalizability

It is important to note the limitations of the generalizability of this sample. It is reasonable to claim that the sample is adequately representative of secondary math specialist and elementary generalist students enrolled in the program, and may also be cautiously interpreted as representative of secondary math specialist applicants who do not later enrol (either as they were unsuccessful in their applications, or they declined offers of admission). Generalizability to applicants and students at other academic institutions must be thoughtfully and cautiously considered, as it is unknown whether applicants to the MT Program can be considered as representing the same population as prospective math teachers at large. Of particular note is the relatively unique nature of the program, and its commonly described "inquiry approach" which serves to significantly differentiate it from teacher education programs at other Western Canadian universities.

Description of the Raw Data

As mentioned previously, the source of the data was the application documents kept on file by the Division of Teacher Preparation at the University of Calgary. These documents consist primarily of academic transcripts from all post-secondary institutions attended by individual applicants. Various internal documents regarding application details and status were also included. The files themselves are kept in storage, organized alphabetically by year of admission. Files are labelled with the individual's choice of program visible (e.g. secondary mathematics, elementary generalist, etc.). Retrieval was a simple matter of pulling out all files from the two programs of interest. All available secondary math specialist files were retrieved and recorded, and all elementary generalist files were retrieved, but only every other record was recorded. In this way, an adequate sample size was obtained, with minimal opportunity for sampling error.

Academic transcripts

Academic transcripts were used to identify all post-secondary degrees awarded to the individual. Recorded information included the name of the institution granting the degree, the type of the degree itself (i.e. Bachelor, Master, or Doctorate), the major or specialization of the degree, and the year in which the degree was awarded.

All mathematics and statistics, and related coursework, receiving a grade of "D" or better was recorded (in accordance with the program admission policy – a "D" is valued at 1.0 on a 4-point scale). Repeated courses were recorded only once, with the highest grade awarded (again, in accordance with the admission policy).

Mathematics and statistics courses were defined as any of the courses offered by the University of Calgary's Department of Mathematics and Statistics, and any similar courses offered by other institutions. This includes, but is not limited to courses entitled Math, Applied Math, Pure Math, Statistics, and Actuarial Science. Related courses were broadly defined as any of those courses offered within the University of Calgary's Faculty of Science, not including those from the Department of Mathematics and Statistics. The related courses include but are not limited to Biological Sciences, Chemistry, Computer Science, Geology, Physics and Astronomy, Environmental Science, Natural Sciences, and Earth Sciences. Again, similar courses from other institutions were included in this category. This categorization was favoured over a more finely detailed course-by-course categorization because of the potential for a serious lack of reliability in the latter. Such detailed judgement would also require significant knowledge regarding each individual course syllabus, which was simply not reasonable in this study.

Detailed records regarding the mathematics and statistics coursework were included for obvious reasons. The "related" coursework was included as there is a theoretical argument for the relevance of science knowledge to the teaching of mathematics. The disciplines of science often provide opportunities to apply mathematics in concrete settings, and to contextualize certain mathematical relations and properties. An interaction between mathematics and science knowledge for science teachers was established by D. H. Monk in his study of the subject area preparation of math and science teachers: "To the degree an understanding of science presupposes some mathematical sophistication, the presence of both science and mathematics course work in a teacher's background could be viewed as an indirect measure of the depth of the subject matter knowledge being offered by the teacher" (Monk, 1994, p. 130). While he did not perform the same investigation for math teachers, he did not discount the possibility. The inclusion of science coursework in this study reflects the possibility and likelihood that some knowledge of science will be relevant for math teachers.

Special attention was given to courses from the faculties of Medicine and Engineering, and courses offered by other faculties entitled "Quantitative Methods". All courses offered by the Faculty of Medicine were included as related courses, as they appear to have much more in common with courses offered by the Faculty of Science than any other faculty. Most courses from the Faculty of Engineering were also included as related with the following exceptions: any courses entitled "Probability and Statistics" (e.g. Engineering 319 at the University of Calgary) or "Numerical Methods" (e.g. Engineering 407 at the University of Calgary) or similar were included as mathematics and statistics courses; courses relating to the professional and legal responsibilities of engineers (e.g. Engineering 513 at the University of Calgary) were not included as either mathematics and statistics or related courses. "Quantitative methods" courses offered by other faculties (defined as any course including "quantitative methods" in the title) were included as "related" courses. There were very few "quantitative methods" courses taken by individuals in the study.

Many records contained partial academic transcripts or none at all. Where the transcript record was noted to be incomplete, the total number of courses that were included was recorded. Finally, there were many cases where the applicant was registered in courses scheduled to conclude well after the application to the program was made. As such, there was no grade awarded at the time of evaluation. It was assumed, however, that applicants would successfully receive credit for these courses, as they appeared to almost always be necessary for the completion of degree requirements. These grades were recorded as "credit received", and were not included in the average GPA calculation.

Other application documents

Other internal application documents were recorded where available and appropriate. The data of interest from these other documents included the following: the gender of the individual; the type of application made (traditional, diverse qualifications, or aboriginal application policy – this determined the admission criteria to which each individual was subject); and the status of their application (e.g. whether the application was complete, if the individual declined an offer of admission, etc.).

The admission evaluation for the program includes the calculation of the average grade point average for the last ten half-courses (where a full Bachelor degree is comprised of 40 half-courses) taken by the individual prior to application to the program. Both undergraduate and graduate courses were treated equally, in accordance with the admission policy. Almost all files specified this average GPA on internal documents, and this was recorded where available. Additionally, previous study of a portion of the secondary math specialist population had shown that the content area and level (i.e. first-year introductory course vs. advanced senior-level course) of these last ten courses were of particular interest (Pascuzzo, 2007, p. 14-21); therefore, the titles and grades received for each of these courses was recorded for the secondary math specialist population only.

Depending on the format of the files (which varied over the years of interest), many of the applicants who had graduated from the University of Calgary also had their final high school grades in mathematics listed. These were recorded where available. The high school courses included Pure Math 30, Applied Math 30, and Math 31 (Calculus).

Finally, notes were made regarding special conditions and situations, and information that was not known to be of definite importance. For example, this included applicants who had graduated from grade 13 systems (as was the case in Quebec and Ontario), or those who had applied in previous years.

Anonymity and Security

As this study makes use of highly sensitive documents and information, every effort was made to maintain the anonymity and security of individual records. The application files themselves were retrieved from the Division of Teacher Preparation storage system, transcribing of data was conducted in available space within the same office, and files were immediately returned to the storage system. No personally identifying information was recorded: each individual record was assigned a unique and arbitrary identification number for indexing purposes only (i.e., a four to seven digit code in which prefix numbers identify the year in which the application was made, followed by an arbitrary index number).

The level of detail afforded by the information recorded could still permit the identification of certain individuals. For example, there is a very small number of applicants presenting previous degrees from certain universities, and that knowledge coupled with the year of graduation could potentially be sufficient to identify some individuals. Similarly, very few applicants arrived at the program with advanced degrees of any kind, and knowledge of an individual's major area would likely be sufficient to enable individual identification based on the information recorded. This is a serious concern, so information that could potentially allow the identification of an individual is not reported, and individual records are kept in secure storage. A very small number of sample records that are determined to be sufficiently "average" so as to prevent individual identification are given to aid in interpreting the data. Findings are only reported at the aggregate level, where the unit of analysis contains five individuals at the very least.

Initial Recording of Data

The aforementioned details were recorded directly into a Microsoft Works file on a laptop computer within the Division of Teacher Preparation office. A complete record included the following information:

- arbitrary unique identification number
- gender
- high school math grades
- previous degree
    - university granting the degree
    - type of degree
    - major or specialization area
    - year graduating
    - additional degrees awarded
- application type
- admission status
- GPA calculated for admission evaluation
- transcript status
- coursework record for all mathematics and statistics and related courses
    - course name (including full course title for mathematics and statistics courses from institutions other than the University of Calgary)
    - grade awarded
- *Secondary mathematics specialists only* - coursework record for the last ten courses taken prior to application (this includes courses from all subject areas, not just those identified above)
    - course name (as above)
    - grade awarded

See Appendix A for samples of original and transcribed records.

Coding of Data

Once retrieval and recording of all records was completed, the records were transferred from Microsoft Works to Microsoft Excel. At this point, coding of the application type, admission status, transcript status, previous degree information, and coursework information was undertaken. All of these variables were coded in a similar fashion: comprehensive lists of values for each variable were compiled using basic Excel functions (i.e. cut, paste, sort, and count). While laborious, this allowed a thorough investigation of all possible options, and permitted recognition of statistically common values and important and practically significant uncommon values. For example, there were 52 different post-secondary institutions granting previous degrees in the combined samples, with as few as one applicant from the particular institution or as many as 248. Without close analysis of uncommon values, however, it would not have been observed that more applicants within the sample have completed previous degrees at institutions in Nova Scotia than in British Columbia (20 vs. 16, respectively). Similar variation was observed with degree type (level and major), coursework, and application type and status.

Admission variables

The Division of Teacher Preparation permits applications for admission under three distinct policies: traditional admission, diverse qualifications, and aboriginal admission policy. In traditional admission, applicants are evaluated on the basis of their academic performance in the ten half-courses completed immediately prior to application. There is a minimum threshold of 2.50 for the average GPA for the last ten courses, and applications are evaluated on a competitive basis given a predetermined quota for each program stream. Under the diverse qualifications policy, applicants are not evaluated competitively based on the GPA of their last ten courses, though they are expected to meet the minimum GPA threshold of 2.50. Individuals applying under this policy must demonstrate that they have achieved excellence in non-academic areas, or have high potential in academic or research activities, or have persevered under great difficulty or hardship (http://www.ucalgary.ca/admissions/diverse_qualifications, November 12, 2007). This must be done by submitting a personal profile with the relevant information, and providing two letters of reference from respected members of the community who can support the information in the individual's profile. The aboriginal admission policy is similar to the diverse qualifications admission policy, but available exclusively for Aboriginal students. Codes were assigned to application type as follows: "0" was used for traditional admission; "1" was used for diverse qualifications; and "2" was used for the aboriginal admission policy.

Admission status proved to be a surprisingly complicated variable: Applicants who were offered admission to the program were coded as "1"; those whose applications were incomplete (e.g. did not submit all transcripts) were coded as "2"; those who were deemed to have deficient content or whose degree was not recognized were coded as "3"; those who did not meet the minimum GPA threshold of 2.50 were coded as "4"; those who did not satisfy TOEFL requirements were coded as "5"; those who cancelled their applications were coded as "6"; those who did not receive accreditation from an external organization in time (e.g. Alberta Learning evaluation of foreign degrees) were coded as "7"; those who met the minimum GPA threshold, but did not meet the quota under traditional admission policy were coded as "8"; those who were not evaluated as they accepted offers in other streams within the MT Program were coded as "9".

Transcript status was more simply coded: files with no transcript records were coded as "0"; complete transcripts were coded as "1"; partial transcripts were coded as "2", and complete 30-course transcripts (as in the case of Ontario and Quebec's grade 13 graduates) were coded as "3".

Institution granting previous degree

As mentioned previously, the comprehensive list of the institutions granting degrees to applicants in the combined samples proved to be sizable. Any institution granting degrees to five or more applicants in the combined samples was given consideration for forming a unique code (this represents approximately 1% of the total of 489 degrees observed within the combined sample). Eleven codes were used to classify the institutions. They are listed below, each with the total number of degrees awarded to individuals within the combined samples:

1. University of Calgary - 248
2. University of Lethbridge - 18

3. University of Alberta - 10
4. Other institution in Alberta - 13
5. British Columbia institution - 13
6. Saskatchewan or Manitoba institution - 11
7. Ontario institution - 40
8. Quebec institution - 7
9. Newfoundland and Labrador, Prince Edward Island, Nova Scotia, or New Brunswick institution - 31
10. International institution - 29
99. Unknown institution - 63

After close inspection of the comprehensive list of institutions granting degrees, it was determined that a semi-geographical approach to the coding would be appropriate. More resolution was desirable within the province of Alberta, as several institutions accounted for a large majority of degrees granted. Less was needed for both the other prairie provinces and the Atlantic provinces, though it should be noted that 20 of the 29 degrees awarded by institutions in the Atlantic provinces were granted by universities in Nova Scotia.

Type of degree

The types of post-secondary degrees were found to conform to the common classification of Bachelor, Master, and Doctorate. One additional category was necessary in order to identify the Combined Bachelor of Education degree offered by the University of Calgary. Students in this program are expected to study three years within a traditional subject area (e.g. mathematics, Canadian Studies, etc.), followed by entry into the two-year after-degree Bachelor of Education program. Upon completion of the B.Ed. program, students are awarded two Bachelor degrees: one in their subject area, and one in education. Previous study of the secondary math specialist group showed that students from the combined program differed significantly from those holding "traditional" bachelor degrees (Pascuzzo, 2007, p. 10), and are thus coded as a distinct group.

The type of degree variable was coded as follows: "0" for no degree or incomplete degree; "1" for Bachelor degree; "2" for the University of Calgary's combined degree; "3" for Master degree; "4" for doctorate; and "99" for unspecified. There were a very small number of advanced degrees represented in the combined sample (15 Master and two Doctorate). It was determined useful to keep the Doctorate category separate, as they represent a very unique portion of the population.

Degree major

Coding of the degree major was conducted with reference to the University of Calgary's structure of divisions, departments, and faculties. Most transcripts stated the major area of the degree awarded as part of the degree title. In several cases, it was not stated, but could be clearly determined by close analysis of the individual's coursework (e.g. more than 20 courses in a particular area). In many other cases, it could not be clearly determined, and was recorded as unknown.

Again, a comprehensive list of all major areas was compiled, and any major reported five times or more was considered for unique coding status. It was determined that a greater number of codes than apparently necessary would be desirable at this point, as codes can easily be collapsed later, but cannot be split once they are established. Thus, a total of 12 codes were used for the degree major variable, and are detailed with the number of corresponding degrees as follows:

1. Mathematics and Statistics – 93
2. Other Science – 20
3. Engineering – 24
4. Computer Science – 5
5. General Studies – 29
6. Psychology – 27
7. Other Social Science – 89
8. Fine Arts – 17
9. Kinesiology – 18
10. Education – 4
11. Other – 25
99. Unknown – 138

The categories include all major programs offered by the corresponding division, department, or faculty at the University of Calgary. Degree majors awarded by other institutions corresponded with these categories very well. "Other Science" includes all major programs offered within the Faculty of Science at the University of Calgary that were not already accounted for by the "Mathematics and Statistics" and "Computer Science" categories. The "Other Social Science" code includes a great number of major areas from the faculties of social sciences, humanities, and communication and culture (as per the faculty divisions specified by the University of Calgary). "General Studies" and "Psychology" were found to be large enough to each merit a unique code from the "Other Social Science" category. The next largest major from within this category was sociology, which represented only nine degrees from the combined sample, so it and other (less frequent) majors were combined in the "Other Social Science" category. The "Other" code represents a broad range of otherwise uncoded majors, including but not limited to management, social work, and law. The major codes of "Computer Science" and "Education" each account for a relatively small number of degrees awarded to applicants in the combined sample, but both are thought to represent groups of sufficient uniqueness and particular interest so as to merit their own codes. Again, they can easily be combined with another code later if it is found to be prudent, but they cannot be split from a code once they are added.

Degrees with double majors were found to be fairly unusual; when they were observed, they were coded by whichever major was judged to be most "mathematics-like" in nature. In practice, and without exception, this happened to be a very clear determination. For example, if the two majors listed were Statistics and Sociology, it was coded as "Mathematics and Statistics" (a clear distinction within the coding scheme). Alternately, if the two majors were History and Geography, it was coded as "Other Social Science" (an immaterial distinction within the coding scheme).

Coursework – "Mathematics and Statistics" courses

The coursework recorded for each individual was already sorted into "Mathematics and Statistics" and "Related" categories (as detailed previously). A comprehensive list of every course taken by each individual within the elementary generalist and the secondary specialist samples was compiled and sorted. Two lists were thus established and kept separate initially, to ensure that any differences in coursework between the two samples could be observed. This proved to be unnecessary, as the categories developed for the secondary math specialist sample were suitable for application to the elementary generalist sample.

An attempt was made to code the mathematics and statistics courses according to the Alberta Program of Studies for K-12 Mathematics. This document identifies four core strands within the mathematics curriculum: number, patterns and relations, shape and space, and statistics and probability. These categories were not observed to align well with the coursework records from the samples, though their utility may be well explored in further post-hoc analyses, and the lack of alignment is expected given the observations of Shulman (1986) and Usiskin (2001). As an example of the problems associated with using codes identified in the program of studies, consider introductory calculus, the mathematics course topic most frequently observed in the combined sample. It encompasses three of the four core strands from the program of studies: number (both concepts and operations), patterns and relations (variables and functions), and shape and space (two and three dimensional objects). While the program of studies categorization would be more immediately relevant to the types of mathematics that prospective teachers are expected to teach, the results of applying such a coding scheme would be quite difficult to interpret given the amount of overlap and the necessity of multiple codes for many of the most frequently observed courses.

All mathematics and statistics course names were recorded so that full details of their content could be obtained. Courses offered by the University of Calgary were recorded as abbreviated course titles (e.g. MATH 249), while courses from other institutions had their full course titles recorded (e.g. Mathematics 1000 - Calculus 1). This way, the courses from the University of Calgary could now be identified in full by referencing university documents, and equivalency of courses could be established. For example it was found that Mathematics 249, Mathematics 251, Applied Mathematics 207, and Applied Mathematics 217 (all offered by the University of Calgary) were equivalent to a first-year introductory course in calculus, or "Calculus 1" from other institutions (the course numbers vary over time, and the Applied Mathematics courses were designed for engineering students). Thus, equivalencies for all courses were established, and accurate tallies for all types of courses were compiled. For example, 103 individuals from the secondary sample had a "Calculus 1" course in their record, and 12 from the same sample had an advanced "Number Theory" course.

Using the compiled list of equivalencies, repeats of topics were identified and eliminated from the coursework records. That is, if it was found that an individual took a course at one institution, and later took an equivalent course at another institution that uses a different course title, the course record with the lowest grade received was removed from the record.

It was apparent that a distinction should be made between courses offered in the first year of post-secondary study, and those offered later. Thus, courses were coded as "j – junior", or first year, "s – senior", or 2nd year and later undergraduate courses, or "g – graduate", or graduate-level courses. While the number of "graduate" level courses was very small, they were thought to be potentially very important, and could easily be collapsed into the "senior" level at a later time if necessary. This "course level" code would be the first part of a two-part code for each individual course: the second being the specific content covered.

Any course type that was taken by five or more individuals was considered for unique coding status. It was quite clear that five codes would capture a very large majority of the topics represented by the coursework in the combined sample. It was also observed that there were a small but likely practically significant number of individuals who had taken a course entitled "Mathematics Appreciation" or something similar; this course type was given a unique code because it is one of the few courses recommended to applicants to the elementary generalist program. Coursework codes were established as follows:

1. Calculus – "c"
2. Linear Algebra – "l"
3. Pure Mathematics – "p"
4. Applied Mathematics – "a"

5. Statistics and Probability – "s"
6. Mathematics Appreciation – "e"
7. Other Math – "o"

In applying these codes, the "Other Math" category contained just slightly more than 5% of the total number of mathematics and statistics courses represented in the combined sample; this appeared to be an adequate level of resolution for this classification. This "course content" code was the second of the two-part code for each individual course in the records. The full coursework code would thus be "jl" for a junior linear algebra course, or "ss" for a senior statistics and probability course.

Coursework grades recorded as percentage scores or on a 9-point scale were converted to a 4-point scale using the same standard grade conversion standards that the program uses (most grades were already in the 4-point scale format). Grades were retained uncoded as a numerical score.

Coursework – "Related" courses

A similar process was used to identify categories for the "Related" courses. However, relatively broad classifications were justified here (because the "Related" courses were not to be central to the analysis), so only abbreviated course titles were recorded (e.g. BIOL 231, ASTR 305, etc.). A comprehensive list of "Related" courses from each sample revealed a fairly simple structure: courses could rather easily be classified within five categories:

1. Physical Sciences – "rp"
2. Life Sciences – "rl"
3. Computer Science – "rc"
4. Engineering – "re"
5. Other Applications – "ra"

These "related" course codes were again the second of a two-part code: for example, a junior level physical sciences course is coded as "jrp", while a graduate level computer science course is coded as "grc".

The distinction between "Physical Sciences" and "Life Sciences" was made in an effort to reduce the size of an "All Sciences" (excluding Computer Science) category. It was determined that this was a reasonable categorization, as this division is popularly recognized. Thus, any science course involving some aspect of the study of living things was placed in the "Life Sciences" category. This stands in contrast to "Physical Sciences", which includes but is not limited to astronomy, physics, chemistry, and geology. The "Computer Science" and "Engineering" categories should be self explanatory. It should again be noted that two Engineering courses (Numerical Methods, and Probability and Statistics) were coded as "Mathematics and Statistics" courses, and other engineering courses relating to the professional and legal responsibilities of engineers (e.g. Engineering 513 at the University of Calgary) were not included.

The final code, "Other Applications", includes courses that do not fit in the first four categories, yet clearly involve the application of mathematics and statistics. This category primarily includes accounting and quantitative research methods. These courses are not considered under the "Mathematics and Statistics" categories, as the emphasis is on the application of mathematical and statistical concepts.

Coursework – "Last Ten" courses

A simplified coding scheme was established for the "Last Ten" courses, as the comparisons to be made for that purpose were to be more broad. An identical "course level" categorization was used (junior, senior, graduate), but the "course content" codes were simplified to three categories:

1. Mathematics and Statistics – "m"
2. Related – "r"
3. Other – "o"

Here, the first two categories reflect the initial broad categorizations, and the third, "Other" category includes any other course. These would not be, by definition, offered by the Faculty of Science, nor would they include an application of any mathematical or statistical concepts.

Coding process

Coding was conducted manually, variable by variable. That is, first, all university names were coded, followed by all degree types, followed by all degree majors, and so on. This way, consistency and accuracy was best ensured. Please see Appendix A for a sample coded record.

First-order Transformation of the Data

Tallies were compiled for numbers of courses in each specified category within individual records. That is, the number of each of the seven types of "Mathematics and Statistics" courses over the three "levels" (21 such categories: e.g. junior calculus, senior applied math, etc.) and the five types of "Related" courses over three "levels" (15 such categories: e.g. junior life sciences, senior engineering, etc.) was obtained using Excel's "count" function. Average grades were obtained for the two broad categories of "Mathematics and Statistics" and "Related" courses using Excel's "average" function (this calculates the arithmetic mean of the values).

The "last ten" courses (secondary math specialist records only) were treated with more detail: tallies and average grades were obtained using the same "count" and "average" functions over each of the available categories. Thus, there is a tally and an average grade listed for each of the nine course variables (three "levels" by three "content" specifications: e.g. junior mathematics and statistics, senior related, graduate other, etc.). Please see Appendix A for a sample transformed record.

Transferral to SPSS

Records were copied from Excel and pasted into SPSS Version 15.0 (Statistical Package for the Social Sciences). Variables were defined and labeled, and data parameters and definitions were established according to guidelines from a reference manual (Norusis, 2005). All records were initially placed into one SPSS file (from both the elementary and secondary samples), from which additional files were created in order to facilitate analyses within subgroups (e.g. all elementary, all secondary, all enrolled in program, etc.).

Formation of Variables

Certain additional variables were created using simplified coding schema, in order to better facilitate analysis, comparison, and interpretation. These include the university granting previous degrees (additional variable created with only two codes: "University of Calgary" and "Other Institution"), admission status (additional variable created with only two codes: "offered admission", and "not offered admission").

Several composite variables were created, again to better facilitate analysis, comparison, and interpretation. For example, the categories for numbers of courses taken within each category (previously classified into 21 mathematics and statistics categories, and 15 related categories) were collapsed first by "level" (i.e. junior, senior, and graduate level) to give seven mathematics and statistics, and five related categories, and also by broad discipline to give one value for the number of mathematics and statistics courses, and another for related courses. Recursive referencing to original data files was used to ensure accuracy for each simplified and composite variable. Other composite variables included the number of courses at each course level (without reference to discipline), and the number by type and level of courses taken within the last ten, with corresponding average grades.

Descriptive Analysis

Analyses were conducted upon variables over several different groupings of the samples. That is, some analyses were conducted upon the combined samples (N=465), with others conducted on individual samples (e.g. elementary, N=289; secondary, N=176), and some upon subsamples (e.g. secondary – offered admission, N=133; secondary – not offered admission, N=43). Within each respective sample or subsample, the number of applicable records was recorded where appropriate. In each case, the method of "pairwise deletion" was applied, rather than "listwise deletion", so as to retain the maximum sample size for each individual analysis.

Variables composed of nominal data (e.g. gender, University granting previous degree, etc.) were analyzed and represented using frequency values for each category. Variables composed of continuous or ordinal data (e.g. high school mathematics grade, number of mathematics courses, etc.) were analyzed and represented using the arithmetic mean and standard deviation for each category.

Significance Testing

In the interest of economy of space, only instructive or insightful comparisons between samples and subsamples over what were determined to be appropriate variables were conducted. These determinations were based on theoretical justification. For example, it was thought to be useful to compare the average high school grades of the elementary and secondary samples, as this might help understand some of the differences between the groups. It was not determined to be instructive however, to compare the average number of mathematics and related courses taken between the elementary and secondary samples, as individuals in the secondary sample are required to have completed a full degree in mathematics or a related discipline in order to be considered for admission. Thus, individuals in the secondary sample will have necessarily enrolled in a much higher number of mathematics and related courses, and a test of the difference in average number of such courses taken between the secondary and elementary samples will reflect this well-understood difference.

Establishing significance

Statistically significant differences between samples and subsamples were identified using common single-variable measures of significance: chi-squared and t-test. These tests were conducted using SPSS 15.0 and according to procedures identified in a reference manual (Norusis, 2005). The null hypothesis for each test was "There are no differences between the tested populations over the given variable(s)". Statistical significance was determined by observing likelihood of differences in distribution or mean values being less than or equal to 5% for both types of significance test (two-tailed distribution in the case of the t-test, as there was no theoretical reason to suspect a directional difference as per Borg & Gall, 1989, p. 550). Using this threshold both satisfies convention and is appropriate to the given situation (Hopkins, Glass, & Hopkins, 1987, p. 142). The ".05 significance level" may be interpreted to mean that one in every twenty findings of significance will be due to type I error, that is, a "false positive" due to random chance (Borg & Gall, 1989, p. 549).

Within those differences that were found to be statistically significant, attention was given to identifying practical significance. Hopkins, Glass, and Hopkins (1987, p. 170-171) state that statistical significance is a necessary but not sufficient condition for practical significance: large sample sizes can permit statistical significance to be observed from differences of trivial or dubious practical importance. As such, each finding of statistical significance is also interpreted in light of establishing practical significance. However, the converse differs: The lack of statistical significance in an observed difference (given that all assumptions for the test are valid) implies that the difference may simply be due to random chance, and that the differences could disappear, or occur with different magnitude or direction if the test were repeated on a different sample from the same population. This does not constitute "proof" of the null hypothesis (that there are no significant differences between the populations), but rather that the data is consistent with the null hypothesis. This is analogous to a declaration of "not guilty" as opposed to "innocent".

It is known, however, that this can result in type-II error, or a "false negative" finding. In attempting to address these errors, differences that were found to be significant at the 0.10 level were also identified separately. These differences deserve some attention as it may be more appropriate to use this significance level for some variables especially in exploratory studies (Borg & Gall, 1989, p. 358), and they may also be indicative of type-II error (Cohen et al, 2005, p. 197). They are included for the sake of discussion, and may be important areas for close analysis in future research.

The next two sections detail the two types of significance testing that were used: the chi-square test, and the independent samples t-test. It is not intended to be an in-depth presentation of the theory and practice of each, but rather a brief introduction that will help to provide some context and interpretability to the procedures used in the study.

The chi-square test

The chi-square test was used to establish whether or not the observed distribution of the tested populations within categorical variables was significant. That is, it shows if a significant relationship between categorical variables exists (Norusis, 2005, p. 446).

For example, consider a sample of 100 university students for which we also know the major for each student (broadly classified as "science" and "arts and humanities"). 50 students are male, and 50 are female. If there is no relationship between gender and major, then we would expect exactly 25 males to be "science" majors and 25 males to be "arts and humanities" majors (and likewise 25 each for the females). A perfectly equal distribution is unlikely in reality, however: while a 26 male "science" / 24 male "arts and humanities" with a corresponding 24 female "science" / 26 female "arts and humanities" distribution may simply be statistical noise, greater deviation from an even distribution will become statistically significant at some point. The chi-square statistic gives a measure of the aggregate deviation of each observed value from the expected values for a particular distribution. If the value of chi-square is sufficiently high (this determination is made by considering the number of categories and combinations of variables possible, or "degrees of freedom"), one can establish the existence of significant relationships between categorical variables with a given level of certainty.

To continue the example from above, consider the observed and expected values given in the following contingency table:

Table 1 - ***Frequency Distribution of Gender and Major (Example)***

Gender	Major		Total
	Science	Arts / Humanities	
Male - Observed	30	20	50
Male - Expected	25	25	
Female - Observed	20	30	50
Female - Expected	25	25	
Total	50	50	100

The chi-square value for the observed distribution is calculated to be 4, which corresponds to a probability of 0.046 given the number of categories and combinations possible. As this probability is less than the chosen threshold value of 0.05, it can be concluded that the observed distribution significantly differs from the expected distribution. Therefore, this data demonstrates that there is a relationship between the variables of gender and major. In this case, we can see that males are more likely to major in "science" than "arts and humanities", while the opposite is true for females.

The chi-square test is a widely used procedure that is relatively easy to interpret, and has few restrictions. The only *a priori* restriction is that categories must be mutually exclusive within the variables compared (but not between them, as dependence between variables is the object of the test). Previously, it was thought that expected values must exceed five in each cell, but it has been shown that the chi-square test functions well with average expected frequencies as low as two (Glass, Hopkins, & Glass, 1987, p. 193). However, this study adopts a general convention of avoiding low cell values (less than five) simply to aid in the interpretability of the findings.

Findings using the chi-square test show both observed and expected values, but only those distributions found to be significant include the calculated chi-square value, degrees of freedom, and corresponding probability.

The independent samples t-test

The independent samples t-test was used to establish whether or not the mean value for a given continuous variable was significantly different between two samples. The t-test provides a measure that indicates the degree of difference between the observed means in reference to an estimate of the standard deviation of the difference (i.e. how large the difference is, given how variable the difference is estimated to be).

For example, consider a sample of 100 high school students. 50 students are enrolled in "school A", and 50 are enrolled in "school B". All of these students take a particular test, and the average (arithmetic mean) test scores for students from school A is 65% and 70% for students from school B. It is not immediately certain that the 5% difference in mean scores is due to anything other than random variation. In using the independent samples t-test, the observed 5% difference in mean scores is examined in light of the relative sample sizes and the variability of individual scores within each sample. As sample sizes increase, we can be more certain that the observed difference is significant (and likewise, as the magnitude of the difference increases); however, as the variability increases, we are *less* certain of the significance of the difference.

If the standard deviation of the test scores in each group is 10%, then the corresponding t-value is calculated to be 2.5, which corresponds to a probability of 0.0141 (referring to a probability table for t-values). This indicates statistical significance (i.e., the difference is not due purely to chance), as it is less than the 0.05 threshold value used in this study. If the standard deviation of the test scores in each group is 13%, then the corresponding t-value is 1.92, which gives a probability of 0.0578 (and is not statistically significant). Similarly, if the difference in means were slightly smaller at 3.5%, with the original 10% standard deviation in test scores for both samples, the difference is not statistically significant.

There are three assumptions necessary for the application of this type of t-test, and they are described here as per Borg and Gall (1989, p. 548). First is the assumption of independent samples, or that any individual can only belong to one of the samples being compared. This assumption was clearly satisfied in the dataset, as individuals were entered only once in the dataset. They could not, for example, belong to the "elementary" and "secondary" samples simultaneously (similarly, they could not be represented in the "offered admission" and "not offered admission" groups).

The second assumption is that the data is normally distributed over a given variable. Certain portions of the dataset are very likely not normally distributed (e.g. the number of math courses taken by elementary generalists), so the details of this assumption were examined closely. According to Norusis, however, if the sum of the sample sizes for compared groups exceeds 40, the assumption of normalcy is not required (2005, p. 138-139). This was the case for almost all t-tests conducted, and those few that did not meet this condition were found to be not significant. Additionally, Hopkins, Glass, and Hopkins state that the violation of the assumption of normalcy has almost no practical consequences (1987, p. 166), and this is confirmed by Borg and Gall (1989, p. 548). In the end, the t-test was determined to be satisfactory, and was preferred to other tests not requiring the assumption of normalcy (most notably the Mann-Whitney U test) as it is more easily interpretable and its limitations are both more widely understood and less problematic for this application (Borg & Gall, 1989, p. 564-565).

The third assumption is of the equality of variances between the samples over the given variable. Levene's test for the equality of variances was used, and t-tests without the assumption of equal variances were conducted where Levene's test indicated such was the case.

Findings using the t-test report t-values, degrees of freedom, corresponding probabilities, and 95% confidence intervals for the difference of the means. These data are reported in the analysis chapter only for comparisons that were found to be significant, and those approaching significance.

Reporting of Findings

Data is reported in a format that attempts to provide a maximum of both detail and interpretability. Thus, categories that were established in the original coding scheme may have been collapsed depending on the context of a particular observation or comparison. Each instance is noted in the analysis. In general, all results that were found to constitute statistical significance are detailed, while others that approach statistical significance are also identified as such, and included on the basis of their potential interest to the reader. Many such marginal findings reflect the necessity of using relatively small samples for certain comparisons, and might indicate areas for further research using larger sample sizes.

CHAPTER FOUR: FINDINGS

Overview

This chapter is divided into three broad sections. First, there is a descriptive report of demographic information regarding both the elementary and secondary samples. This includes information regarding gender, previous degree, grades in high school mathematics, and grade point average as calculated for the admission criteria. Next, there is a detailed analysis of the findings regarding the coursework of the elementary sample. Finally, there is an analysis of the findings for the coursework and grades received for the secondary sample, including both the "offered admission" and the "not offered admission" subsamples.

Significance test statistics are reported in this chapter for all findings of significance at the .05 level, though attention is also drawn to findings at the .10 level, as mentioned previously.

Demographics

There were two samples included in the study. The first is composed of applicants who were admitted to the elementary generalist stream of the program from 2005 to 2007. The second is composed of applicants who were admitted to the secondary mathematics specialist stream from 2000 to 2007, and applicants who applied but were not admitted to the secondary mathematics stream from 2006 and 2007. Some of these individuals were offered admission, but did not enrol in the program. In order to provide maximum clarity and interpretability, demographic information is reported with two overlapping divisions of the samples. The first comparison details those individuals who were enrolled in the program (both elementary and secondary, and indicated in table titles using "Enrolled Students"); the second details the secondary subsamples not on the basis of enrolment, but rather whether they were offered admission or not (indicated by "Secondary Applicants"). It was determined that most direct comparison between the elementary and secondary samples should be made with all individuals enrolled in the program, but that the important distinction at the secondary level was that of whether or not the individual was offered admission.

Gender

As is well understood to be the case, there were more females than males in both the elementary and secondary samples. As Table 2 shows, the difference was less pronounced in the secondary sample, and a chi-square test confirmed that gender and level of program (i.e. elementary or secondary) were not independent ($\chi^2 = 25.2$, $df = 1$, $p < .0005$). While the expected gender distribution in the elementary sample was 68.4 males and 219.6 females, there were 49 and 239 observed, respectively. Conversely, the expected distribution in the secondary sample was 27.6 males and 88.4 females, but 47 and 69 were observed. Thus, there was an over representation of females in the elementary sample, and an over representation of males in the secondary sample. The observed gender distribution amongst the subsamples of successful and unsuccessful secondary applicants was not observed to be significant, which is consistent with the independence of gender and success in application.

Table 2 - *Gender Distribution for Enrolled Students and Secondary Applicants*

	Frequency (%)			
Sample	Male	Female	Missing	Total
		Enrolled Students		
Elementary	49 (17.0)	239 (82.7)	1 (0.3)	289 (100)
Secondary	47 (40.5)	69 (59.5)	0 (0.0)	116 (100)
Total	96 (23.7)	308 (76.0)	1 (0.2)	405 (100)
		Secondary Applicants		
Offered Admission	52 (39.1)	81 (60.9)	0 (0.0)	133 (100)
Not Offered	17 (39.5)	22 (51.2)	4 (9.3)	43 (100)
Total	69 (39.2)	103 (58.5)	4 (2.3)	176 (100)

Previous degree

Three aspects of individuals' previous degrees were recorded: location, major, and level. In the case of individuals holding multiple degrees, the degree most closely related to mathematics was recorded as the primary degree. If this distinction was still unclear, the most recently awarded degree was recorded. There were very few individuals in the combined sample holding multiple or advanced degrees. Consequently, detailed analyses concerning the degree level data were not conducted.

The location of institutions granting previous degrees was coded into ten geographically delineated categories. Table 3 shows the distribution of location of previous degree over the elementary and secondary samples of enrolled students: A chi-square test showed that the distribution was significant ($\chi^2 = 18.7$, $df = 9$, $p = .03$). However, there were numerous expected values less than 5 (35% of cells), so the finding must be interpreted cautiously. Additionally, the difference between observed and expected values does not appear to be generally practically significant, with the possible exception of an observed over representation of degrees from the Atlantic provinces in the elementary sample, and over representation of degrees from international locations in the secondary sample. Location of previous degree was unknown for 57 individuals in the elementary sample and 3 from the secondary sample.

There is an apparent contradiction between Table 3 and Table 4: Table 3 lists a total of 113 secondary applicants for whom the location of their previous degree was known, while Table 4 gives a total of 170. The difference arises because Table 3 lists only secondary applicants who were also enrolled in the program (as this is the basis of comparison between the elementary and secondary groups), while Table 4 shows all applicants to the secondary stream. Of the 130 secondary applicants who were offered admission, 17 declined, giving the net value of 113 enrolled secondary applicants listed in Table 3.

Table 3 - *Location of Previous Institution of Enrolled Students*

Location	Elementary Frequency (%) Observed	Elementary Frequency (%) Expected	Secondary Frequency (%) Observed	Secondary Frequency (%) Expected	Total
Calgary	141 (60.8)	143.9 (62.0)	73 (64.6)	70.1 (62.0)	214
Lethbridge	12 (5.2)	8.7 (3.8)	1 (0.9)	4.3 (3.8)	13
Edmonton	3 (1.3)	3.4 (1.4)	2 (1.8)	1.6 (1.4)	5
Other Alberta	11 (4.7)	8.1 (3.5)	1 (0.9)	3.9 (3.5)	12
Other Prairie	5 (2.2)	5.4 (2.3)	3 (2.7)	2.6 (2.3)	8
BC	12 (5.2)	11.4 (4.9)	5 (4.4)	5.6 (4.9)	17
Ontario	16 (6.9)	19.5 (8.4)	13 (11.5)	9.5 (8.4)	29
Quebec	4 (1.7)	2.7 (1.2)	0 (0.0)	1.3 (1.2)	4
Atlantic	21 (9.1)	17.5 (7.5)	5 (4.4)	8.5 (7.5)	26
International	7 (3.0)	11.4 (4.9)	10 (8.8)	5.6 (4.9)	17
Total	232 (100)	232 (100)	113 (100)	113 (100)	345

The analysis of location of previous degree for the secondary applicants used a simplified location code in order to avoid many low expected low cell values. Two categories were used: degrees granted by the University of Calgary, and those granted by other institutions. The distribution shown in Table 4 was not found to be significant, which is consistent with the location of previous degree being independent of success in application.

Table 4 - *Location of Previous Degree of Secondary Applicants*

Location	Offered Admission Frequency (%) Observed	Offered Admission Frequency (%) Expected	Not Offered Frequency (%) Observed	Not Offered Frequency (%) Expected	Total
Calgary	79 (60.8)	74.9 (57.6)	19 (47.5)	23.1 (57.6)	98
Other Institution	51 (39.2)	55.1 (42.4)	21 (52.5)	16.9 (42.4)	72
Total	130 (100)		40 (100)		170

The major area of previous degree was analyzed using 11 categories, and is shown in Table 5. Differences in distribution between the elementary and secondary samples of enrolled students confirmed the expectation that those in the secondary specialist stream would be more likely to hold degrees in mathematics and related disciplines. A chi-square test demonstrated the statistical significance of this distribution ($\chi^2 = 236.4$, $df = 10$, $p < .005$). There were low expected values for science, computer science, and education majors (27.3% of cells). Information regarding major area of previous degree was missing for 105 individuals from the elementary sample, and 14 from the secondary sample.

Table 5 - *Major Area of Enrolled Students*

	Elementary Frequency (%)		Secondary Frequency (%)		
Major	Observed	Expected	Observed	Expected	Total
Math / stats	1 (0.5)	41.8 (22.7)	64 (62.7)	23.2 (22.7)	65
Science	5 (2.7)	8.4 (4.5)	8 (7.8)	4.6 (4.5)	13
Engineering	0 (0)	11.6 (6.3)	18 (17.6)	6.4 (6.3)	18
Computer Science	0 (0)	2.6 (1.4)	4 (3.9)	1.4 (1.3)	4
General Studies	27 (14.7)	18.0 (9.8)	1 (1.0)	10.0 (9.8)	28
Psychology	25 (13.6)	17.4 (9.4)	2 (2.0)	9.6 (9.4)	27
Other Social Science / Humanities	76 (41.3)	50.8 (27.6)	3 (2.9)	28.2 (27.6)	79
Fine Arts	14 (7.6)	9.0 (4.9)	0 (0)	5.0 (4.9)	14
Kinesiology	17 (9.2)	11.6 (6.3)	1 (1.0)	6.4 (6.3)	18
Education	2 (1.1)	1.9 (1.0)	1 (1.0)	1.1 (1.0)	3
Other	17 (9.2)	10.9 (5.9)	0 (0)	6.1 (5.9)	17
Total	184 (100)	184 (100)	102 (100)	102 (100)	286

Major area of previous degree for secondary applicants was analyzed using a simplified coding scheme with three categories, again to avoid many low expected values. Table 6 shows the distribution for secondary stream applicants who were successful and unsuccessful in their applications. A chi-square test demonstrated that major and success in application are not independent ($\chi^2 = 9.2$, $df = 2$, $p = .01$). This is unsurprising, given that individuals not holding previous degrees in mathematics or a related discipline are less likely to have met the minimum requirements for admission. The distribution in Table 6 confirms the intuition that this effect would be most pronounced with mathematics and statistics majors, and less so with those holding majors in related disciplines (applicants with "Other Science" majors are observed to be less successful than their counterparts holding "Math / Stats" majors).

Table 6 - ***Major Area (Simplified) of Secondary Applicants***

	Offered Admission Frequency (%)		Not Offered Admission Frequency (%)		
Major	Observed	Expected	Observed	Expected	Total
Math / Stats	73 (62.9)	70.2 (60.5)	16 (51.6)	18.8 (60.5)	89
Other Science	33 (28.4)	30.8 (26.5)	6 (19.4)	8.2 (26.5)	39
Other non-Science	10 (8.6)	15.0 (12.9)	9 (29.0)	4.0 (12.9)	19
Total	116 (100)	116 (100)	31 (100)	31 (100)	147

High school mathematics grades

Information regarding high school mathematics grades was available for a relatively small subsample of both the elementary and secondary groups. These individuals had all received previous degrees from the University of Calgary. It was unknown if there was any systematic difference between individuals listing high school grades in their records and those not doing so, as only a portion of individuals holding degrees from the University of Calgary had this information. Thus, these analyses must be interpreted cautiously. The high school courses listed are those that are currently offered in local schools; historical equivalents were included for the purposes of these analyses (Math 33 is now Applied Math 30; Math 30 is now Pure Math 30).

Table 7 shows the high school grades received for elementary and secondary students in the program. No individuals in the secondary sample listed Applied Math 30 grades. The difference in average grade received for Pure Math 30 was found to be statistically significant via independent samples t-test ($t = 7.89$, $df=82.4$, $p < .0005$). In this case, Levene's test for the equality of variances failed; the reported test statistics were calculated to accommodate this. The mean difference in Pure Math 30 grades between the two groups was 14.5%, an observation of certain practical significance. A 95% confidence interval for the mean difference was found to be 10.9% to 18.2%. The mean difference in Math 31 (Calculus) grades was not significant.

Table 7 - *High School Mathematics Grade for Enrolled Students*

	Applied Math 30		Pure Math 30		Math 31	
Standard	Elementary	Secondary	Elementary	Secondary	Elementary	Secondary
Mean	70.6	n/a	68.8	83.3	79.5	80.8
SD	8.5	n/a	10.4	7.5	7.1	9.8
n	8	0	66	32	12	24

Significant differences in high school grades between successful and unsuccessful secondary applicants were evident. Differences in the Pure 30 grades failed to meet the .05 level of significance: A mean difference of 4.4% was observed with $p = .075$. A 95% confidence interval was found to span from -0.456 to 9.233. Differences in Math 31 grades were not found to be noteworthy.

Elementary Sample

Not all records contained full transcript information. Table 8 summarizes the number and type of transcripts available within the elementary sample. Full coursework analyses were restricted to the subsample of elementary records containing full transcript information from 40-course degrees. The decision to exclude other records was carefully considered. Records with no transcript information could not be used for this type of analysis. Individuals with full 30-course degrees would appear to have a lower number of courses taken in every category, and were a sufficiently small portion of the sample that their exclusion could be considered reasonable.

Partial records could be considered useful, as one could use relative measures of course content (e.g. math courses as a percentage of all courses taken) instead of absolute measures (e.g. total number of math courses taken). This was not reasonable, however, as it was observed during data collection that non-math majors (i.e. most elementary students) tended to take math courses at the beginning of their academic program, if at all. Partial records contained course information in reverse chronological order, and as such were consistently missing data from the early part of individuals' degrees. Thus, inclusion of the partial records would have most likely skewed the representation of the sample by minimizing the already low number of mathematics courses taken by elementary generalists. Partial records were excluded, as it was determined that the cost of their inclusion outweighed the benefit. It was unknown if the subsample with full 40-course records was representative of those with partial records, but there was no reason to suspect that there was a systematic factor at work. Nevertheless, the full 40-course subsample was compared with the partial transcript subsample over the available variables to identify if significant differences existed. As expected, the group with partial transcripts had significantly fewer courses in every category, but differences in grades received for courses were not significant. Additionally, GPA as calculated for the admissions criteria (average of the last ten courses taken, based on a 40-course degree) was not found to be significant. Previous degree major and institution granting previous degree was significant, however: General Studies majors, University of Calgary graduates, and females were over represented in the partial transcript subsample, while males and psychology and other social science / humanities majors were over represented in the full 40-course transcript group.

Table 8 - ***Elementary Student Transcript Completion Summary***

Transcript Type	Frequency	Percent
None	19	6.6
Partial	114	39.4
Full 30-course	4	1.4
Full 40-course	152	52.6
Total	289	100

Coursework

Student coursework was represented by tabulating the frequency and type of courses taken. Table 9 shows the frequency and percentage of the elementary full 40-course transcript subsample of individuals having taken a specific number of courses of two types. Please note that the rows of the table have uninterpretable sums. Each column, however, adds to the sample size of 152 individuals, or 100%. While 86 individuals (56.6% of the subsample) did not take a single mathematics and/or statistics course, 35 (23.0%) did not take a related course. Twenty-nine individuals (19.1%) did not take a mathematics / statistics or related course. The table illustrates the tremendous variety of coursework experiences: The distribution is bimodal (not including the "Math and Stats" column), though the second mode is much smaller (comprised of individuals having taken 13 or more math and related courses.

Table 9 - ***Number of Math and Related Courses for Elementary Students (n = 152)***

	Frequency			Percent		
Number of Courses	Math and Stats	Related	Math/Stats and Related	Math and Stats	Related	Math/Stats and Related
None	86	35	29	56.6	23.0	19.1
1	29	22	21	19.1	14.5	13.8
2	20	25	18	13.2	16.4	11.8
3	9	22	21	5.9	14.5	13.8
4	2	12	17	1.3	7.9	11.2
5	0	3	5	0.0	2.0	3.3
6	5	7	6	3.3	4.6	3.9
7	0	3	4	0.0	2.0	2.6
8	1	2	5	0.7	1.3	3.3
9	0	4	5	0.0	2.6	3.3
10	0	2	0	0.0	1.3	0.0
11	0	3	3	0.0	2.0	2.0
12	0	3	0	0.0	2.0	0.0
13	0	0	5	0.0	0.0	3.3
14	0	2	2	0.0	1.3	1.3
15 or more	0	7	11	0.0	4.6	7.2

The number and type of the mathematics courses taken by these same individuals are detailed in Table 10. While most individuals did not take any mathematics or statistics courses, those that did were most likely to take one or two courses in calculus or probability and statistics. The percentage of individuals having taken at least one course for each of the categories is as follows: calculus – 24.4%; linear algebra – 7.2%; probability and statistics – 19.7%; pure math – 0%; applied math – 0.7%; math appreciation – 9.8%; other math – 5.3%.

Table 10 - ***Number and Type of Math and Stats Courses for Elementary Students (n = 152)***

Number of Courses	Frequency						
	Calculus	Linear Algebra	Probability and Statistics	Pure Math	Applied Math	Math Appreciation	Other Math
None	115	141	122	152	151	137	144
1	24	6	18	0	1	11	7
2	10	5	11	0	0	4	1
3	3	0	0	0	0	0	0
4	0	0	1	0	0	0	0

Similar detail is given for the number and type of related courses taken by elementary students. Table 11 shows that courses in the physical and life sciences were most frequently taken, though a relatively great number of individuals took one course in computer science. The percentage of individuals having taken one or more courses in each category is as follows: physical sciences – 51.3%; life sciences – 48.0%; computer science – 41.4%; engineering – 0%; other related – 5.9%. There is not such a strong case for a bimodal distribution here, but the cluster of individuals taking eight or more life sciences courses might be considered a unique subsample, and serves to further differentiate between this distribution and that of the mathematics and statistics courses.

Table 11 - ***Number and Type of Related Courses for Elementary Students (n = 152)***

Number of Courses	Frequency Physical Sciences	Life Sciences	Computer Sciences	Engineering	Other Related
None	74	79	89	152	143
1	32	22	56	0	5
2	16	22	4	0	2
3	10	9	3	0	1
4	5	5	0	0	0
5	4	3	0	0	0
6	5	3	0	0	0
7	3	1	0	0	0
8	0	2	0	0	1
9	0	2	0	0	0
10 or more	3	4	0	0	0

Grade point average

The average grades received for math and stats courses and related courses are shown in Table 12. The average grades received for both categories were very similar, and both were about 0.6 grade points lower than the average used in the admission criteria. In fact, the "admission" GPA exceeded the "math / stats" and "related" GPA for all but 21 individuals in the elementary subsample. Note that the variability in the average GPA is much more pronounced for the mathematics and related courses than for the admission GPA.

Table 12 - ***Average GPA for Elementary Students with 40-course records***

Course Type	Mean	*SD*	*n*
Math and Stats	2.67	0.76	61
Related	2.70	0.66	114
Combined Math and Related	2.68	0.66	117
Admission GPA	3.38	0.27	145

Secondary Sample

Again, not all records contained full transcript information. Table 13 summarizes the number and type of transcripts available for the secondary sample. As mentioned in the previous chapter, coursework information for enrolled secondary students was recorded only for individuals who completed a previous degree at the University of Calgary. These were all complete 40-course records. The coursework information for all applicants who did not later enroll (both those found inadmissible and those who declined offers of admission) was recorded, with only a few partial or 30-course records. Again, partial and full 30-course records were excluded from coursework analyses as their benefit did not outweigh their cost. Thirteen of the records with no transcript information were those of graduates of the University of Calgary.

Table 13 - ***Secondary Applicant Transcript Completion Summary***

Transcript Type	Frequency	Percent
None	67	38.1
Partial	3	1.7
Full 30-course	2	1.1
Full 40-course	104	59.1
Total	176	100

Given the nature of the sample of enrolled secondary students (all former graduates of the University of Calgary), the representativeness for the students from other universities was in question. Comparisons were made over the available variables (gender, admission GPA, major, and institution granting previous degree), and significant differences in major area of previous degree were observed between students from the University of Calgary, and those from other institutions. Table 14 shows that there was an over representation of "Math and Stats" majors from the University of Calgary, with an under representation of "Other Science" and "Other non-Science" majors. A chi-square test indicated that the distribution was significant ($\chi^2 = 8.00$, $df=2$, $p = .018$). While this illustrates an important difference within the subsample, it does not suggest that students *with the same major* will differ significantly on the basis of attending a different institution previously.

Table 14 - ***Institution and Major for Previous Degree of Successful Applicants***

	Institution				
	University of Calgary Frequency (%)		Other Institution Frequency (%)		
Major	Observed	Expected	Observed	Expected	Total
Math and Stats	48 (72.7)	41.4 (62.7)	16 (44.4)	22.6 (62.7)	73
Other Science	14 (21.2)	19.4 (29.4)	16 (44.4)	10.6 (29.4)	33
Other non-Science	4 (6.1)	5.2 (7.8)	4 (11.1)	2.8 (7.8)	10
Total	66 (100)	66 (100)	45 (100)	45 (100)	116

Coursework

The average number of courses taken by the admissible and inadmissible secondary applicants is shown in Table 15. While the differences were not found to be significant, the variance in the inadmissible group is noticeably larger than that of the admissible group.

Table 15 - ***Average Number of Courses Taken by Secondary Applicants***

	Admissible (n = 73)		Inadmissible (n = 31)	
Course Type	Mean	*SD*	Mean	*SD*
Math and Stats	14.2	6.0	12.7	6.9
Related	10.7	10.2	12.5	11.1
Math/Stats and Related	24.9	9.8	25.3	13.3

Table 16 details the number and type of mathematics courses taken by successful secondary applicants (those who were offered admission). Again, calculus appears to be the most enrolled topic, but there is a greater number and variety of courses taken by individuals in this sample. The percentage of individuals having taken one or more courses in each topic is: calculus – 98.6%; linear algebra – 93.2%; probability and statistics – 91.8%; pure mathematics – 74.0%; applied mathematics – 78.1%; math appreciation – 34.2%; other math – 58.9%.

Table 16 - ***Number and Type of Math Courses for Admissible Applicants (n = 73)***

	Frequency						
Number of Courses	Calculus	Linear Algebra	Probability and Statistics	Pure Math	Applied Math	Math Appreciation	Other Math
None	1	5	6	19	16	48	30
1	1	10	28	9	13	19	36
2	6	55	20	17	15	3	6
3	12	3	5	13	8	0	0
4	52	0	4	6	10	1	0
5	0	0	2	4	7	2	0
6 or more	1	0	8	5	4	0	1

The number and type of courses taken by inadmissible applicants appeared to have a similar distribution. Table 17 shows that calculus and probability and statistics were oft-enrolled courses. The percentage of individuals having taken one or more courses in each topic is: calculus – 93.5%; linear algebra – 80.6%; probability and statistics – 93.5%; pure mathematics – 54.8%; applied mathematics – 51.6%; math appreciation – 54.8%; other math – 41.9%.

Table 17 - ***Number and Type of Math Courses for Inadmissible Applicants (n = 31)***

Number of Courses	Frequency: Calculus	Linear Algebra	Probability and Statistics	Pure Math	Applied Math	Math Appreciation	Other Math
None	2	6	2	14	15	14	18
1	2	7	13	3	2	2	10
2	7	18	9	5	4	4	1
3	4	0	3	2	3	4	0
4	16	0	2	3	3	3	2
5	0	0	0	1	2	2	0
6 or more	0	0	2	3	2	2	0

The average number of each type of math course (as shown in Table 18) reflects the above distribution. Significant differences were found between the admissible and inadmissible groups for average number of calculus ($t = 2.53$, $df = 42.1$, $p = .015$), linear algebra ($t = 2.34$, $df = 46.7$, $p = .024$), and math appreciation courses ($t = -3.32$, $df = 36.8$, $p = .002$). Levene's test for the equality of variances was failed in each case. The mean difference and 95% confidence intervals for each of those types of courses are: calculus – 0.64, 0.13 to 1.14; linear algebra – 0.38, 0.05 to 0.71; math appreciation – -1.27, -2.05 to -0.50. For this latter course type, note that the sign differs: it is the inadmissible applicants who enrol in more of these courses.

Table 18 - ***Average Number of Math Courses Taken by Secondary Applicants***

Course Type	Admissible (n = 73) Mean	Admissible (n = 73) *SD*	Inadmissible (n = 31) Mean	Inadmissible (n = 31) *SD*
Calculus	3.60	0.86	2.97	1.28
Linear Algebra	1.77	0.64	1.39	0.80
Statistics	2.75	3.57	2.23	2.58
Pure Math	2.40	2.85	2.00	2.76
Applied Math	2.38	2.22	1.71	2.04
Math Appreciation	0.53	1.03	1.81	2.02
Other Math	0.75	0.97	0.65	1.05

The number and type of related courses taken by admissible secondary applicants is quite variable. Table 19 shows that there are noticeably more outliers in each category (particularly in engineering, as six individuals had taken 20 or more such courses), and a general tendency toward the physical sciences. The percentages of individuals who have taken one or more courses in each category are: physical sciences – 83.6%; life sciences – 37.0%; computer science – 76.7%; engineering – 24.7%; other related – 6.8%.

Table 19 - ***Number and Type of Related Courses for Admissible Applicants (n = 73)***

	Frequency				
Number of Courses	Physical Sciences	Life Sciences	Computer Sciences	Engineering	Other
None	12	46	17	55	68
1	15	11	29	3	1
2	11	6	17	2	3
3	9	2	1	2	0
4	6	1	2	0	1
5	4	2	0	1	0
6 or more	16	5	7	10	0

The distribution of number and type of related courses taken by inadmissible applicants appears to be slightly less disperse (shown in Table 20). The percentages of individuals having taken one or more courses in each category are: physical sciences – 87.1%; life sciences – 45.2%; computer science – 61.3%; engineering – 22.6%; other related – 12.9%.

Table 20 - ***Number and Type of Related Courses for Inadmissible Applicants (n = 31)***

Number of Courses	Frequency: Physical Sciences	Life Sciences	Computer Sciences	Engineering	Other
None	4	17	12	24	27
1	2	7	9	1	0
2	7	4	2	0	2
3	4	1	4	0	1
4	2	0	0	1	0
5	2	1	2	0	0
6 or more	10	1	2	5	1

The average number of related courses does not appear to differ significantly between the admissible and inadmissible samples. Table 21 shows that while the mean difference between the numbers of physical science courses taken was quite large (3.4 more courses taken by the inadmissible group), it was also associated with a large standard deviation and small sample size. Consequently, Levene's test for the equality of variance was failed, and the mean difference did not achieve statistical significance (t = 1.78, df = 33.5, p = .085). However, a 95% confidence interval for the mean difference ranges from -0.46 to 6.78, so it is quite likely that the mean difference between the populations which would be found to be statistically significant in a larger sample.

Table 21 - ***Average Number of Related Courses Taken by Secondary Applicants***

	Admissible (n = 73)		Inadmissible (n = 31)	
Course Type	Mean	*SD*	Mean	*SD*
Physical Sciences	3.26	3.52	6.42	9.64
Life Sciences	1.42	3.21	1.19	2.64
Computer Sciences	2.27	3.97	2.03	3.58
Engineering	3.59	8.86	2.39	6.02
Other Related	0.15	0.62	0.48	1.57

Last ten courses

The composition of the last ten courses taken by applicants to the secondary stream is given in Table 22. The last ten courses are given under two different classifications. The first places the courses into three categories, arranged by topic. The second classifies the courses according to whether they were junior level (i.e. first-year) or senior / graduate level. Differences between the admissible and inadmissible groups were observed to be small, and none were found to be statistically significant.

Table 22 - ***Average Composition of Last Ten Courses For Secondary Applicants***

	Admissible (n = 68) Number of Courses		Inadmissible (n = 28) Number of Courses	
Course Type	Mean	*SD*	Mean	*SD*
		By Topic		
Math and Stats	3.54	2.39	3.14	2.64
Related	2.07	3.01	2.04	2.70
Other	4.38	2.61	4.82	2.71
		By Level		
Junior	1.85	1.59	1.71	1.68
Senior / Graduate	8.15	1.59	8.29	1.68

Grade point average

Table 23 gives the average grades received by secondary applicants for overall coursework and the last ten courses. Overall coursework grades are given for mathematics and statistics courses, related courses, and a weighted average of math / stats and related courses. The last ten courses grades are again divided by topic and level, but also include a weighted average of math / stats and related courses. As one would expect, the differences in averages between the admissible and inadmissible groups are statistically significant (as the criteria for admission are primarily academic competition). Each average was found to be significant, except for the junior level courses taken in the last ten. It is interesting to note that mean differences ranged from 0.236 for overall related coursework to 0.505 for last ten mathematics and statistics. Mean differences were generally greater for the average grades of the last ten courses.

Table 23 - *Average GPA for Secondary Applicants*

	Admissible			Inadmissible		
Course Type	Mean	*SD*	*n*	Mean	*SD*	*n*
			All Coursework			
Math and Stats	2.93	0.48	66	2.64	0.34	27
Related	2.84	0.56	62	2.60	0.46	27
Math/Stats and Related	2.91	0.43	64	2.63	0.32	27
			Last Ten Courses			
Math and Stats	3.02	0.55	56	2.52	0.67	19
Related	3.15	0.57	36	2.69	0.68	14
Other	3.36	0.48	61	2.94	0.59	25
			Last Ten Courses – Alternate Categories			
Math/Stats and Related	3.07	0.50	64	2.64	0.58	24
Junior	3.45	0.41	53	3.15	0.75	19
Senior / Graduate	3.17	0.41	65	2.80	0.52	26

Table 24 shows a series of second-order variables that were computed from simple operations on the variables listed in Table 23. Note that these values cannot be obtained directly from those given in the previous table, as the mean difference was computed on an individual basis, and the sample sizes are lower due to pairwise deletion over the two variables considered. The first series of variables was obtained by subtracting individual "overall math and stats" grades from the "last ten math and stats", "overall related" from "last ten related", and "overall weighted math / stats and related" from "last ten weighted math / stats and related". None of these differences between the admissible and the inadmissible groups were found to be statistically significant. The second series of variables was formed by subtracting each of "overall math and stats", "overall related", and "overall weighted math / stats and related" from "admission GPA".

Only the last of these variables was found to be statistically significant (overall weighted math / stats and related; $t = 2.06$, $df = 87$, $p = .042$), though the others may be worth closer analysis (overall math and stats; $t = 1.76$, $df = 89$, $p = .081$; overall related; $t = 1.96$, $df = 85$, $p = .054$). 95% confidence intervals for each mean difference consistently indicate that it is likely that the admissible group has greater such differentials (lower bounds of the mean difference of -0.02, -0.004, and 0.006 respectively).

Table 24 - ***Average Differences Between Overall and Last Ten GPA***

	Admissible			Inadmissible		
Comparison	Mean Difference	*SD*	*n*	Mean Difference	*SD*	*n*
			"Last Ten" Minus "Overall"			
Math	0.083	0.320	56	-0.080	0.525	19
Related	0.202	0.513	36	-0.050	0.413	14
Math/Related	0.169	0.382	62	0.021	0.426	24
			"Admission GPA" Minus "Overall"			
Math	0.327	0.400	66	0.150	0.489	25
Related	0.404	0.470	62	0.174	0.557	25
Math/Related	0.334	0.346	64	0.157	0.413	25

CHAPTER FIVE: DISCUSSION AND CONCLUSION

Overview

The discussion is presented in four parts. First, there is a summary of the findings which engages in some interpretation with respect to the context of the study. Second, there is a more detailed interpretation of key findings, together with potential implications. Third, a discussion of the limitations of the study, and how said limitations should be considered in interpreting the findings. Finally, there is a section detailing suggestions for future research, both in analyzing the existing data and in performing new research that will help to qualify, interpret, and establish the generalizability of the findings.

Summary of Findings

Selected findings are detailed for both descriptive data and determinations of significance. This section is divided into three main parts: demographic findings, elementary sample – specific, and secondary sample – specific.

Demographics

The gender distribution of the elementary and secondary samples is of little surprise. It is important to note that there is no evidence that an individual's gender is related to their success in application to the secondary mathematics specialist program. As noted in the methodology section, this does not prove that there is no such relationship, but rather that the data is consistent with there being no relationship. This is important information for contextualizing the differences that do exist between the admissible and inadmissible subsamples.

The location of institution granting previous degree was found to be related to whether an individual was enrolled in the elementary or secondary stream of the program. While numerous low cell values hamper the interpretation (35% of cells with a value of 5 or less), there were notable practical differences observed: an over representation of elementary students from institutions in the Atlantic category (and possibly other non – University of Calgary Albertan institutions), and an over representation of secondary students in the Ontario and International categories. It is interesting to note that a large majority of individuals holding previous degrees in the Atlantic category graduated from institutions in Nova Scotia (20 out of 29 such individuals in the combined sample). It is also interesting to note that a larger portion of the combined sample hold degrees from institutions in Nova Scotia than from institutions in British Columbia, a close neighbour. Comparisons for the same variable over the secondary applicant groups required a simplified coding scheme: University of Calgary, and other institution. Location of previous degree and success in application to the secondary program was not observed to be significant, again important in contextualizing the differences between the groups.

Rather unsurprisingly, there are proportionately far more mathematics majors in the secondary sample. While the preponderance of social science and humanities majors in the elementary sample is no surprise, it is interesting to see that the two most commonly identified majors in the area are general studies and psychology. Within the sample of secondary applicants, the distribution of major confirms that individuals holding math / stats degrees are more likely to be offered admission to the program than their counterparts with unrelated degrees.

The data regarding high school mathematics grades must be interpreted carefully, as it is unknown if there is systematic bias in the reporting of these data. No individuals in the secondary sample reported having taken Applied Math 30. This is likely because this course is not widely accepted by universities as a prerequisite for entry into mathematics or related degree programs. Within the elementary sample, students received slightly higher grades in the Applied Math 30 course than in the Pure Math 30 course (70.6% and 68.8%, respectively), though very few reported having taken Applied Math 30. Students in the secondary sample received significantly higher grades in Pure Math 30 (83.3%) than those in the elementary sample (68.8%), with observably lower variation within their scores. Students in the secondary sample are likely to have received grades from just over 10 to almost 20 percent higher than their counterparts in the elementary sample (as per a 95% confidence interval). That students in the secondary sample are significantly more successful in their final high school mathematics course should not be a surprise, but the magnitude of the difference is noteworthy. Differences in grades received for Math 31, Calculus, are negligible, though they are based on very small sample sizes. The admissible and inadmissible secondary groups show marginal difference in their high school grades: a mean difference of 4.4% in favour of the admissible applicants was observed, with marginal statistical significance.

Elementary sample

Analysis of the coursework of the elementary sample is limited to those individuals who had full transcript records for 40-course degrees. These findings are purely descriptive, as there are no subsamples to compare. Of those included, it is noted that 86 individuals (56.6% of the sample) had not taken a single mathematics or statistics course, and 29 had not taken a math / stats or related course (19.1% of the sample). About half of the sample took from one to four math / stats and related courses (77 individuals, 50.7%), while a small but noticeable portion took 13 or more math / stats and related courses (18 individuals, 11.8%). This last group should not be regarded as outliers, as they likely represent those individuals in the sample who hold a major in a math or related discipline (or those who began such a program, but later changed), and form a unique portion of the sample. The distribution is also notable in that more individuals in the sample had not taken a single math / stats or related course than had taken 10 or more courses.

The table regarding the type of mathematics courses taken shows that calculus and probability and statistics are the two most popular types of math courses in the elementary sample. One or more calculus courses were taken by 24.4% of the sample, with 19.7% for statistics and probability. These topics are likely over represented here because many individuals take calculus in their first year of study, as it is a requirement for degree programs in science. Unfortunately, a relatively small number of individuals reported having taken a mathematics appreciation course similar to the Math 205 course offered at the University of Calgary (less than 10% of the sample).

The coursework detail for related topics is more variable. One or more physical science courses were taken by 51.3% of individuals, and though slightly fewer took one or more life science courses (48.0%), there were more who took eight or more life science courses. Of particular note here is that 41.4% of individuals reported taking one or more computer science courses, though almost all of these had taken only one such course. This may be due to the fact that an introduction to computers course was the most frequently reported course in the entire elementary sample (76 individuals took Computer Science 203 at the University of Calgary). Incidentally, the next two most frequently reported courses in the entire elementary sample were Geology 209 (46 individuals), and Astronomy 205 (33 individuals). This refers only to the courses offered at the University of Calgary, and does not include equivalent courses at other institutions. That these three courses were the most frequently reported is not surprising, as they are widely regarded to be "easy" courses offered by the Faculty of Science for non-science majors, and they are popular undergraduate courses for this reason. Unofficially, the respective names of these courses are "Bits for Twits", "Rocks for Jocks", and "Scopes for Dopes", and the University of Calgary student newspaper regularly recommends them as "easy" courses ("Classes: What to Take," 2006, p. 9); a cursory internet search reveals that these unofficial designations are not unique to the University of Calgary.

The average overall grades received for both math / stats and related courses taken by individuals in the elementary is approximately a "B-", with a standard deviation of about 0.7 grade points. The average grade received in the last ten courses prior to application to the program for these individuals is just over a "B+" with a standard deviation of 0.27 grade points. Thus, grades received in math / stats and related courses are both lower and more variable than grades received in the last ten courses. It may be argued that individuals tend to receive lower grades in courses taken outside of their major area, and that the incentive of admission inspires individuals to achieve higher grades for the last ten courses taken. The lower variability noted in the GPA of the last ten courses is likely due to the threshold effect of the GPA-based admission criteria: The mean GPA for all three years is 3.38, but the distribution is truncated slightly below the mean once the quota for admission is reached. This threshold for admission is slightly different from year to year, with an average of 3.01. This results in a positively skewed distribution of the average GPA for the last ten courses, with a median value of 3.34 (slightly less than the mean). The average grades received for all math / stats and related courses have a more normal distribution.

Secondary sample - coursework

The representativeness of the portion of the enrolled secondary student sample with coursework information is limited. All of these individuals hold previous degrees from the University of Calgary, and it was shown that location of previous degree has a statistically significant relationship with the major of previous degree. More math and stats majors are observed to have graduated from the University of Calgary, with a corresponding under representation of "other science" majors. Thus, the secondary sample used in the coursework observations has a bias in favour of math and stats majors. However, there is no theoretical reason or empirical evidence to suggest that individuals holding previous degrees with similar majors are significantly different based on the institution granting the previous degree. While this is not impossible, there is nothing to suggest it, so the aforementioned bias in the secondary – coursework sample is all that concerns us here.

The average number of math / stats and related courses taken by admissible and inadmissible applicant samples do not differ significantly. The variability in the inadmissible sample appears to be greater than that of the admissible sample, however.

The nature of the math and stats courses taken appears to have a very different distribution than that for the elementary sample. Again, calculus is the most frequently reported course topic (98.6% of individuals in the sample took one or more of these courses), with most applicants (71.2%) having taken four calculus courses. This is consistent with the fact that four calculus courses form a core component of most degree programs in mathematics. Linear algebra is the next most frequently reported course topic, with 93.2% of individuals taking one or more courses. Just as calculus forms a core component of most degree programs in math, so does linear algebra: 75.3% took two courses in linear algebra, which is a common program requirement. None of the other course categories exhibit the strong modality evident in the calculus and linear algebra categories. Probability and statistics is the next most frequently reported type of course, with 91.8% of individuals taking one or more of these courses. Pure and applied mathematics are commonly reported (74.0% and 78.1% took one or more courses, respectively) as well. There seems to be a number of outliers in the probability and stats, pure math, and applied math categories: each show at least nine individuals taking five or more courses in the area. This may be consistent with individuals holding degree majors in each of those areas.

The coursework information for the inadmissible secondary applicants is similar in that the strong representation of and modality within the calculus and linear algebra categories is repeated. However, there is generally a less math-intensive coursework profile evident here: almost half of the sample had not taken any courses in either pure math or applied math.

In direct comparison, the mean differences between number of courses in each area for admissible and inadmissible groups showed several statistical differences. The individuals in the inadmissible sample were less likely to take calculus and linear algebra courses, but more likely to take a math appreciation course.

The related coursework for admissible applicants is heavily skewed towards the physical sciences and computer science. There is a noticeable collection of outliers in each course category, possibly indicating degree majors in each of those areas. Related coursework for the inadmissible applicants has a similar general distribution, but is less variable, with fewer outliers in each course category except for the physical sciences (and perhaps engineering).

Direct comparison between the admissible and inadmissible groups for the average number of courses taken in each category shows only one mean difference of potential significance: number of physical science courses. A 95% confidence interval describes the mean difference as between -0.46 to 6.78 courses. The likelihood of a large difference here is great, but not certain. The large variability could be attributed to a relatively small portion of the inadmissible sample taking a great number of physical science courses and skewing the group average.

Secondary sample – last ten courses

The precise composition of the last ten courses taken by individuals prior to their application to the program is of particular interest. Applicants are aware that the admissions process is competitive, and that average grade received for the last ten courses prior to application is the criterion for admission (excluding those applying under alternate admissions policies). Thus, special attention was drawn to their choice of courses and corresponding grades received for this important period.

Observed differences in the composition of last ten courses between the admissible and inadmissible groups were both practically and statistically not significant. This comparison was made first by comparing course composition by topic (math, related, and other categories), then by level (junior and senior/graduate categories).

Secondary sample – grades

Average grades received were calculated for all math and stats courses, all related courses, a weighted average for all math / stats and related courses, and all categories of courses taken in the last ten. For each type of average grade, the admissible group received higher grades than the inadmissible group. All of the observed differences except that for junior level courses taken in the last ten are observed to be statistically significant. This is consistent with the fact that the admissible group is differentiated from the inadmissible group primarily on the basis of grades received in the last ten courses. The fact that a similar difference was observed when considering all degree coursework supports the notion that the grades received in the last ten courses are representative of overall grades received, at least as far as discerning between the two groups is considered.

An additional series of observations can be made in comparing mean average grades within the admissible and inadmissible groups. It is observed that overall grades received for each category of courses are lower than the corresponding grades for the last ten courses. For example, admissible applicants averaged 2.93 for math and stats courses overall, but 3.02 for math and stats courses taken in the last ten. This is also the case for related courses, and the weighted averages of math / stats and related courses. The case of the inadmissible applicants is similar, but less pronounced. In fact, inadmissible applicants received higher overall average grades in math and stats than in the last ten courses. Additionally, it is the case for both the admissible and inadmissible groups that the average grades received in the last ten courses is lowest for math and stats courses, and highest for courses not related to mathematics. Given that the average composition of the last ten courses taken by both admissible and inadmissible samples includes a large portion of courses unrelated to mathematics and statistics (almost half of the last ten courses: 4.4 and 4.8 courses, respectively), the fact that the corresponding grades for these same courses are significantly great than those for math and stats courses is worthy of attention (0.34 and 0.42 grade points for the admissible and inadmissible samples, respectively).

Key Findings and Implications

There are three areas that receive detailed description and interpretation in this section: the limited amount of the mathematics coursework taken by elementary stream students, the highly variable nature of the coursework taken by secondary stream applicants, and the potential for unintended consequences of the admission criteria for both streams of students.

Elementary coursework

It is well understood by teacher educators that applicants to the elementary generalist program have a variety of subject-area backgrounds, that relatively few have backgrounds in the sciences, and that a very small number have backgrounds in mathematics. This is confirmed in the data, but the degree of under-representation of individuals with science and mathematics backgrounds is perhaps surprising, at just over three percent. One half of one percent of individuals in the sample (in this case, one person out of the subsample of 184) have a background in mathematics.

As teachers in elementary schools are expected to be subject-area generalists, this extremely low representation may be due to degree requirements within mathematics and science. It may be the case that it is simply "easier" to pursue a well-rounded university education (suitable for a generalist educator, perhaps) within another faculty. That is, it is possible that the degree requirements within the faculty of science are more rigid than in other faculties. If that is the case, then the apparent under-representation might not be so troubling. However, an examination of the number and type of courses taken by elementary generalists in their entire degree programs reflects a similar under-representation of mathematics and related subjects: more than half (57%) did not take a single mathematics course, while an additional 19% took only one such course. One wonders what the corresponding values would be for English or social science courses. These and subsequent observations are made in reference to a full undergraduate degree being comprised of 40 courses.

Similarly troubling is the observation that about one quarter (23%) of the individuals did not take a single math-related course (defined as any other non-math course offered within the Faculty of Science). These related courses would provide an opportunity for individuals to at least potentially apply mathematical techniques. Thus, a significant majority of individuals have had no formal education in math since high school, and a significant portion of those have not had any opportunity to apply math in a related area.

While the question of what mathematics knowledge elementary teachers should have is the subject of debate and research, these observations indicate that the formal mathematics education (similar but not equivalent to mathematics knowledge) that a majority of individuals has experienced is limited to their own elementary and secondary education. This is consistent with a common assumption amongst prospective elementary teachers that their pre-university mathematics education is sufficient to teach elementary mathematics (McDiarmid, Ball, & Anderson, 1989, p. 199).

The nature of mathematics courses that were taken is limited as well. More than half (57%) of those who did take a math course (n=66) enrolled in one or more courses in calculus, while slightly less than half (45%) took one or more courses in probability and statistics. These are commonly required courses for degree programs in the sciences and social sciences, respectively (though there were not many science majors in the sample, there were a number of individuals who took a traditional "science program" in their first year of studies, and subsequently changed majors), and probably have a large representation on that basis. There are relatively few individuals (23%) who took a "math appreciation" course designed for teachers (such as Math 205 – Mathematical Explorations offered at the University of Calgary). While it is not the intent of this study to establish the content alignment of post-secondary math courses with the elementary math curriculum (though it is observed to be less than ideal: Goulding, Hatch, & Rodd, 2004), it deserves mention that a far smaller number of prospective teachers took a course designed for math teachers than a course in calculus.

It is also important to note that the average grades received for math and related courses is significantly lower than that for the last ten courses taken. It is unknown if these last ten courses might include a significant portion of math and related courses, but the difference in average grades is striking: While elementary generalists received an average slightly higher than a B+ (3.4) for the last ten courses, their average for math and related courses was much lower at a B- (2.7). This serves to further qualify the finding that even when math courses were taken, the grades received were significantly lower than for the last ten courses, which are most likely composed of non-science courses. That is, prospective elementary teachers tend to take a very small number of math and related courses, and they also tend to enjoy less academic success in these courses than in others.

This area of the findings demonstrates that a majority of prospective elementary teachers have had no formal opportunity to further their learning of mathematics since their own time in high school. If they have, it has been largely limited to introductory calculus, or introductory probability and statistics, and they have been significantly less successful in these courses than in others. Additionally, these courses typically occur in large lecture theatres with several hundred students. The relevance of these topics to the elementary curriculum has not been established, and the context of the lecture theatre may not be conducive to personal interaction with mathematics. This may not be a helpful mathematics experience for prospective elementary teachers who will be faced with the many questions about the foundations and nature of mathematics that elementary students are prone to ask (Goulding, Hatch, & Rodd, 2004).

The specifics of these findings will not likely surprise teacher educators, but the overwhelming general under-representation of mathematics in the post-secondary education of prospective elementary teachers observed here may serve as a call to action. It may be difficult to allocate more time to mathematics instruction in the already full program of study. It may be more reasonable for the MT Program to change the recommendation for taking Math 205 (Mathematical Explorations), or an equivalent, to a requirement. This would be preferable to requiring "any single course in mathematics" for the reasons stated above: the relevance and utility of any particular topic offered for math or science majors is not well-established, whereas a math course designed for teachers has at least a theoretical benefit. As further study illuminates the relationship between teacher mathematical knowledge and classroom practice, more specific recommendations for prerequisite courses or in-program mathematics instruction may be justified.

Secondary coursework

While the pattern of coursework enrolment amongst the prospective elementary teachers can be characterized as uniformly low with few exceptions, the coursework enrolment for prospective secondary teachers is highly variable. Ideally, the analyses would have been conducted on the subsamples represented by the major area of previous degrees obtained: those with previous math degrees, combined degrees, computer science degrees, and engineering degrees (the four largest such groups) all have very distinctive course enrolment patterns, but the limited sample size does not permit meaningful comparison at that level (using the existing data). As such, the data is reported and analyzed at a more macro level, so that observations are made with respect to all applicants to the secondary math specialist stream.

For example, the average number of math and stats courses taken by admissible applicants is 14.2. The average for applicants with prior degrees in mathematics would necessarily be much higher (due to the degree requirements in their respective programs), and the average for applicants with prior degrees in engineering would be much lower. This is reflected in the relatively high values for standard deviation (6.0 courses in this case). The number of related courses is even more variable, with a mean of 10.7 and standard deviation of 10.2. Much of this effect (positive skew) arises from the inclusion of individuals with multiple degrees and prior degrees in engineering in the sample (engineering programs typically require students to take a greater number of courses for the completion of a degree). While these individuals are correctly identified as outliers, their inclusion in these analyses is deemed to be more instructive than their exclusion.

Combining these distinct groups is perhaps the greatest factor contributing to the observed variability. While differences between these groups is of doubtless interest, observations regarding the combined sample are important in understanding the overall mathematics background of applicants to (and hence, the students in) the program. Attention is drawn to the variability in the total sample, but not on the basis of the major area of the applicant's previous degree. The observations regarding coursework will focus on that of the admissible applicants only, as the coursework of the inadmissible applicants was sufficiently similar.

Incidentally, the average number of math and stats courses taken by admissible applicants in this study is significantly higher than that of the practicing teachers sampled in Monk's landmark study of secondary math and science teacher content knowledge (Monk, 1994). He found that the teachers in the LSAY (Longitudinal Survey of American Youth) sample had taken an average of 10.36 math courses, and another 4.72 science courses. The corresponding values for the sample of admissible applicants are 14.2 and 10.7, respectively. The possible explanations for this discrepancy are many, but it is most likely that it occurs because of the fact that the teachers in the LSAY sample would probably have had a great variety of academic backgrounds, including some with degrees in disciplines unrelated to math, and some possessing emergency certification not requiring the completion of an undergraduate degree. In contrast, the sample of admissible applicants is comprised of individuals who have all completed at least one degree in a math or related field.

Mathematics coursework profiles again show a preponderance of calculus courses, with a large majority of admissible applicants (68%) taking four calculus courses. A similar majority (72%) took two linear algebra courses. This is consistent with the fact that there are four core calculus and two core linear algebra courses that are requirements for many degree programs in mathematics and related fields. Greater variability was observed in the taking of pure and applied mathematics courses, as no clear mode was evident in either of these categories. There is a relatively even spread from zero to as many as five of these courses (for each course category). On average, applicants were noted to take at least two courses within each course category identified (this number was slightly lower for linear algebra, given that there is typically fewer of these courses offered). Of particular note here, however, are the large standard deviations for the average number of courses for each course type. For statistics and pure math, the standard deviation exceeds the value of the mean, while the values are very close for applied math.

This type of variability is likely primarily determined by the major area of individuals. Those three topics identified above coincide with three main major areas within mathematics: pure math, applied math, and statistics (including actuarial science according to the coding scheme used in this study). An individual with a major in one of these areas will positively skew the distribution of values, and contribute to a larger standard deviation. The same occurs in observations regarding the related coursework, most notably in engineering as there are 10 individuals having taken 6 or more engineering courses. The mean number of engineering courses for the sample is only 3.6, but the standard deviation is 8.9: not only do engineering students take more courses in a typical degree program, they have fewer opportunities to take courses outside of their area.

This demonstrates the tremendous variability in the coursework profiles of applicants to the secondary stream. The variability is largely due to the major area of the previous degree held by individuals, but individuals are not differentiated on this basis in their application to the program, or after entry into the program. Thus, the former engineering major and the former math major attend the same classes if they have both been accepted into the secondary math specialist stream.

The great variation in coursework profiles is likely a confounding factor in the education production function studies outlined previously. Such studies do not take into account the details of coursework choices and restrictions faced by students as illustrated here. Given that some courses are bound to have different applicability and utility for secondary school teachers than others, they should not be treated as identical. An introductory statistics course (which is taken by more than 90% of secondary applicants) will inform one's understanding of mathematics differently than an advanced course in topology (which is taken by very few secondary applicants), for example. How this might influence classroom practice has not been determined, but if the link between mathematical knowledge and student achievement is to be investigated then it should be acknowledged that mathematical knowledge is domain specific (understanding statistics very well might not help one to effectively teach geometry). Adding this level of detail (the domain specificity of knowledge) to education production function studies might help discern the link between teacher knowledge and effective teaching.

It is again difficult to make specific recommendations as to what particular courses prospective math teachers should take. Regardless of their area of specialization, it is probable that secondary school teachers will not have engaged with the content they are expected to teach (or at least a portion thereof) since their own time as a student in high school (Grossman, Wilson, & Shulman, 1989; Usiskin, 2001; Luk, 2005). If this is the case, further study of how teacher knowledge influences classroom practice may demonstrate how the knowledge gained in advanced study of particular areas of mathematics influences effective teaching.

Currently, the MT Program requires applicants to have completed 10 courses in mathematics, with a minimum of five different sub-domains of math represented. The data present no compelling reason to change this policy, but the variability amongst applicants with regard to their coursework profiles will be important for teacher educators to note. Understanding the prevalence of calculus and linear algebra courses in the coursework of prospective teachers, for example, might be helpful in tailoring the content of the program to better meet the needs of prospective secondary math teachers.

Admission criteria

As was previously mentioned, the composition of the last ten courses taken by secondary applicants is of particular interest. A large majority of individuals apply under the traditional application procedure, and they are likely aware that the process is competitive (on the basis of the grades that they received in the ten courses taken prior to application). As most degree programs allow for a certain amount of latitude in course selection, a closer investigation of the nature of these ten courses is undertaken to determine if and how individuals alter their course selections in order to maximize their competitive advantage in the admissions process. Since there are no restrictions placed on the type of courses that may be taken in this period, it is possible that some individuals may choose to take these last ten courses on the basis of anticipating relatively high grades. The type of coursework choices made will vary widely between individuals: some will be finishing their degree programs and may have to take all ten courses as dictated by their degree requirements, whereas others may have finished their degree requirements and may have the option of choosing any courses that they want. Others, it should be noted, may have been unsuccessful in applying to the program in the previous year, and may have taken courses in the interim in order to "boost their GPA". Regardless, the nature of these last ten courses is unknown, and it is unknown if they are representative of the applicants' overall coursework or of their ability or knowledge in any particular area.

Elementary applicants are expected to be content generalists, so their overall coursework will vary widely (as was shown previously), and one would expect the same of the last ten courses taken in their programs. As such, it may not be reasonable to expect any particular kind of representativeness from those last ten courses. The last ten courses taken by the elementary applicants were not recorded for this reason. The average grade received for the last ten courses was recorded for each individual, however, and this was noted to significantly exceed the average grades received in mathematics and related courses (B+ versus B-, respectively). From an academic success perspective, it may be observed the last ten courses are not representative of coursework in math and related areas.

The nature of the last ten courses taken has greater implications for the secondary applicants, as they are expected to be specialists in their field. The admission criteria state that individuals applying to the secondary math stream of the program should have the equivalent of a major in mathematics or statistics, or at least ten courses in the area (based on a 40-course undergraduate degree). Again, there are no restrictions placed on the type of courses taken in the last ten before application, and thus there are no restrictions placed on the type of courses used in the average GPA calculation for competitive admission. It is possible for an applicant to the secondary math specialist stream to not have taken a single math or related course within the last ten. If this is the case, then the individual would compete for admission on the basis of grades received in courses entirely outside of the realm of mathematics and related disciplines. This might not be noteworthy in the case of elementary generalists, but raises some serious concerns for secondary specialists. If the purpose of the admission competition is to discern on the basis of ability or achievement in a particular subject, then the representativeness of the courses included in the admission competition is critical.

The data shed some light on this question of representativeness. Amongst the admissible secondary applicants for whom the composition of the last ten courses was available (n = 68), the average composition of the last ten courses was 3.5 math courses, 2.1 related courses, and 4.4 other courses. This illustrates that the representativeness of the last ten courses is not ideal for prospective mathematics teachers. A full degree in mathematics or a related science discipline is composed of a large majority of math and related courses, and certainly less than 44% unrelated courses. A very large majority of the secondary sample is comprised of individuals with previous degrees majors in math and related disciplines (approximately 92%), so this over-representation of unrelated courses cannot be considered representative of overall coursework.

A similar analysis of the last ten courses categorized as junior level (first year) or senior / graduate level (second year and beyond) showed the representativeness to be more satisfactory on a level-of-study basis. Secondary applicants take an average of 1.9 courses at the junior level in the last ten courses.

These findings are indicative of the representativeness of average course-taking behaviour. Analysis of individual cases is more revealing, as great variation exists here as well. While no individuals took mathematics courses for all ten of their last ten courses, 15% of individuals took no mathematics courses in the same period. That is, 15% of individuals competed for admission to the secondary mathematics program on the basis of grades received for non-math courses. Fortunately, only 3% of individuals took all ten courses outside of math and related subjects. Similar demonstrations of individual variation are evident in the composition of the last ten courses according to junior and senior / graduate categorizations: While a large majority of individuals took zero, one, or two junior level courses (74%), a small number took five, six, or seven such courses (7%).

This demonstrates that the representativeness of the average type of the last ten courses taken is limited. It also shows that the representativeness is much lower for certain individuals. It may be the case that these individuals take less representative courses in the last ten by chance, but it may also be intentional or tactical course selection in order to maximize their competitive advantage in the admissions competition. An American study of the course selection behaviour of high school students (Finn, Gerber, & Wang, 2002) found that institutional policies and practices were more important than demographic variables (i.e. students selected courses on the basis of the policies and practices to which they were subject). It is reasonable that the same phenomenon may exist at the post-secondary level, indeed anywhere that individuals are subject to "high-stakes" competition.

It was unknown if the observed coursework selection patterns would differ between the admissible and inadmissible groups. Comparisons were made with the inadmissible subsample, and no significant differences were observed for either the average composition of the last ten courses, or the distribution of individual compositions.

The grades received for courses taken in the last ten also serve as an illustration of the representativeness of the admission criteria. The average grades received for math courses taken in the last ten are notably lower than those for unrelated courses (3.02 versus 3.36, respectively), and the average grades received in related courses falls between the two values (3.15). Additionally, the average grade received for junior level courses is higher than that for senior and graduate level courses (3.45 and 3.17, respectively). This indicates that the courses taken in the last ten in which applicants tend to receive the highest average grades are junior level courses, and are unrelated to mathematics. As these types of courses are over-represented in the average composition of the last ten courses, this is consistent with the claim that applicants may engage in tactical course selection for this period.

The differential between average grades for overall coursework and those for the last ten courses is informative. This differential for math and related coursework (average "last ten" math / related course grade minus average overall math / related coursework grade) is 0.169 grade points (based on a four-point scale). That is, admissible applicants received measurably higher grades for math and related courses taken in the last ten than for similar courses taken in their overall program. This observation holds for both math courses and related courses individually, though the differential is notably greater for related courses. The differential between the average grades for overall math coursework and the average grade for all of an individual's last ten courses (the GPA value used in the admission competition) is much greater, approximately one-third of a letter grade (0.327 grade points).

The former differential value compares grades for similar types of courses (math and related courses), and shows that the grades for such courses taken in the last ten are significantly higher than those in the applicants' overall program. It may be likened to an "apples-with-apples" comparison. The second differential value illustrates how representative the average grades received in the last ten courses are of the average grades received for overall math coursework. It may be described as a "fruit-to-apples" comparison (how well does the last ten average grade represent the overall math-only grade). Both differential values show that the average grades received in the last ten courses are not completely similar to the overall grades in math courses.

The differential values are consistently greater for admissible applicants than for inadmissible applicants. The differential is statistically significant for overall math and related courses and the average of all last ten courses, and potentially significant for each of math and related courses and the average of all last ten courses. Even without statistical significance, a 95% confidence interval indicates that the admissible group consistently shows greater such differentials. The general trend (receiving higher grades in the last ten courses) is consistent with a few different scenarios: average grades increase as students spend time in their program; the incentive effect of the admission criteria motivates students to achieve higher grades in the last ten courses; individuals take "easier" courses in the last ten, by chance or by intent.

This significant difference observed between the admissible and inadmissible groups is not readily interpreted. It may be the case that the individuals in the inadmissible group simply "struggle" with math and related subjects throughout their degree program, and cannot improve their grades in the last ten courses in the same way that those in the admissible group do. It may be the case that those in the admissible group simply improve in their respective fields over the course of their degree programs, and the greater differential in grades reflects greater growth. It is also possible that the admissible group is more successful in intentionally maximizing the differential, whether by tactical course selection (by taking "easier" courses, or repeating courses with the expectation of receiving a higher grade), or concerted effort.

This data does not demonstrate beyond doubt that the admission criteria are unfairly biased against individuals on the basis of course selection. The possibility exists, however, that what has been described as tactical course selection could be exploited under the current admissions process. An individual having the ability and means to take a large number of "easy" courses (it is well known that average course grades vary between faculties) in the last ten would have a definite advantage over someone not able to do so (perhaps because of degree requirements, family commitments, or lack of resources). It appears that most applicants engage in some form of tactical course selection (given the general over-representation of unrelated courses in the last ten). It is troubling, however, to speculate that one of the characteristics that serve to differentiate between the admissible and inadmissible groups is the greater success of the former in employing tactical course selection. A well-documented phenomenon in the field of psychometrics is known as "test-wiseness", which includes a range of behaviours that individuals can use to improve their scores that have nothing to do with the object of the test. Thus, the individual's score represents (at least in part) their ability to "beat the test". There is no reason why the same concern should not be considered in the realm of admissions criteria. It is critical that the criteria measure the attribute intended, rather than "system-beating" behaviour.

To this end, it may be appropriate to change the admission criteria. If the purpose is to have applicants compete for admission on the basis of their ability and achievement in their area of specialization, then the criteria should reflect that emphasis. Perhaps the criteria should stipulate "the last ten courses in mathematics and statistics". It may not be the case that *only* math ability is of interest, in which case a balance akin to "the last five math courses and the last five non-math courses" may be appropriate. Alternately, "the *best* five math courses and the last five non-math courses", or some derivative, may be more appropriate. The precise formulation is debatable, but the inclusion of at least some math course grades into the admission competition is desirable.

Limitations

This study, as any, presents a compromise of detail and generalizability (Shulman, 1981, p. 12). Appropriate interpretation within the context of the program studied and generalizability to others requires thoughtful consideration of the unique limitations presented by the data and method of analysis. This section provides a detailed account of limitations and their implications in this context. It also describes apparent limitations that were theoretically or methodologically accounted for in the study, and others that are not applicable given the precise formulation of the data and methods. The limitations are presented in the following order: populations considered, samples obtained, unit of analysis, and methodology.

Population

The first limitation of note relates directly to the parameters of the population. It is not known how similar the applicants to and students in the MT Program are to those of other institutions and programs. While it is undoubtedly the case that many individuals are likely to apply to several such teacher education programs, it is unknown whether there are significant differences between individuals who apply to and/or enrol at different institutions. There is a theoretical argument for the case that applicants to this program ought to be significantly different from applicants to programs at other institutions, as the program actively distinguishes itself as a non-traditional, inquiry-based program. As such, it may be the case that applicants to the MT Program are more philosophically attracted to this program rather than others. It is unknown whether this theoretical difference has a practical effect, or whether it is a significant factor given the tendency for individuals simply to apply to programs on a purely geographic or individual cost basis.

Thus, one cannot be certain that the observations regarding students at and applicants to the MT Program are generalizable to those at other institutions. Only similar studies at other institutions would be able to substantiate such similarity or difference. One can, however, be quite certain of the generalizability of the findings to the distinct populations of enrolled elementary generalist students and secondary math specialist applicants to the program (both enrolled and not enrolled). Naturally, this is limited by the sampling methods used.

Sample

There were two distinct populations identified (enrolled elementary generalist students, and all secondary math specialist applicants), and two different methods used to gather representative samples. The different methods were established to address the unique constraints of the data available for each distinct population.

For the elementary generalist students, this study examines a systematic random sample of enrolled students over three years of the program's operation (2005-2007). This limitation of the population was determined to provide adequate sampling according to established guidelines. Inclusion of students from previous years would not have added to the sample's representativeness for current and future students (these are the individuals of interest in this study).

As mentioned previously, the elementary generalist sample does not include unsuccessful applicants, or those who chose to decline offers of admission. It also does not include elementary French Immersion students (as they were thought to comprise a unique population themselves), or Early Childhood Education specialists. This latter exclusion was necessary because of the logistical difficulty of collecting data for a representative sample of this potentially unique population. It is likely that the observed tendencies of the elementary generalist sample will be somewhat representative of the early childhood education population as well, but the degree of similarity is unknown. This represents an appropriate area for further study.

The secondary math specialist sample encompasses all such applicants to the MT Program from 2000 to 2007. For those enrolled in the program, the sample is equivalent to the population (all records were included in the sample). There are two major constraints regarding the representativeness of this sample: First, the subsample of enrolled students includes coursework information only for students whose previous degree(s) were from the University of Calgary. Second, records for the subsample of applicants who were not later enrolled in the program (including both unsuccessful applicants and those who declined offers of admission) were only available for 2006 and 2007.

The coursework limitation is problematic, as it is unknown whether the coursework of students from other institutions is similar to that of students from the University of Calgary. However, there is no theoretical reason to suspect that it should be different, for those individuals with similar previous degrees. A problem arises in that it was demonstrated that there is a significant relationship between institution granting previous degree and the major of the degree: Individuals with a previous degree from the University of Calgary were more likely to have a mathematics or statistics degree than those from other institutions. Individuals from other institutions were more likely to have a previous degree in a "related" discipline. The secondary coursework data should therefore be carefully interpreted with this in mind (given the over-representation of mathematics and statistics majors in the subsample pertaining to coursework information).

An attempt was made to address the fact that there was limited data available for individuals who were not enrolled in the program (both those deemed inadmissible, and those who declined offers of admission) by including all such records available. Thus, this subsample includes all available information for all applicants to the secondary math specialist stream who were not enrolled in the program (both demographic information and detailed coursework information). While relying solely on two consecutive years of applicants is not ideal, it was the best representation possible, given the limited data available. Again, the findings pertaining to the secondary sample must be carefully interpreted with this in mind.

Unit of analysis

Other limitations arose from the necessity of using existing data that was collected for a different purpose (that of the admissions department). Using data as it already existed was definitively more constraining than developing and applying measures specifically for a given application. The use of the relatively crude measures of type of course taken and grade received illustrates but a small portion of an individual's academic experience. It is, however, an observable and interpretable portion of their experience, and may help to inform a more detailed and in-depth understanding.

Similarly, the findings are reported at an aggregate level that necessarily obscures some of the detail. This was done partially to prevent the inadvertent or intentional identification of individuals from the unique details contained in the records, and partially to better understand the similarities held by individual applicants and students. It is hoped that a suitable compromise of generality and detail has been achieved.

As noted in the literature review, the measures analyzed in this study (number and type of courses, and corresponding grades received) do not comprehensively describe the nature of an individual's experience with mathematics: No single measure could do so. However, the coding and summary of this level of coursework information does describe a well-defined and interpretable portion of that experience, and allows for observations of similarity and difference to be made at the group level. The measures that were used as the object of this study are, like all units of analysis, a compromise of detail and generalizability.

Methodology

This study was conducted in a post-hoc fashion, and as such is arguably subject to greater limitations than studies with a more experimental approach. This has implications for significance testing, specifically detailed below.

The statistical methods used in this study were chosen to be as appropriate and as easily interpretable as possible. That is, the simplest statistical tools were employed, and no oversimplifications of the data were made (such as collapsing categories beyond interpretability, or arbitrarily categorizing ordinal data in order to ease analyses). Attention was given to reporting violations of assumptions where they occurred. These violations mainly consisted of the limitations of representativeness for particular samples and subsamples.

The statistical significance of dependency between nominal variables was established using the chi-square test, as it is a reliable and readily interpretable method. The precise magnitude of significant relationships identified by the chi-square test is not easy to represent. Both coefficient lambda and Goodman and Kruskal's tau were unsatisfactory for this application (given lambda's reliance on modal categories, and tau's reliance on *a priori* knowledge of direction of dependency; Norusis, 2005, p. 174-177). While each of these measures may have been appropriate for one or two individual comparisons, overall consistency in significance testing methods was thought to be more effective. Simple analysis of the cell residual values in the context of relative category size (observed minus expected values, given column and row totals) provides a clear indicator of exactly where the independence model fails (Norusis, 2005, p. 164). It also gives a relative measure of magnitude and direction of a relationship that is easily interpreted and can be applied for all applications of the chi-square test.

Similarly, the independent samples *t*-test is a widely used, reliable, and easily interpretable method for establishing the statistical significance of differences between groups based on ordinal data. The magnitude and direction of relationships identified using the *t*-test are well described by the values of mean difference and 95% confidence intervals of the difference. The confidence interval provides an especially useful measure of the difference, as one can be confident that 95% of mean differences will lie within this interval. As stated in the methodology chapter, the assumptions for the independent samples *t*-test were typically readily met, or deemed to be moderate violations given the relatively large sample sizes (Norusis, 2005, p. 138-139).

The most serious limitations of the methods employed arise from reliance on univariate measures and statistical methods. Cohen et al. state that it is desirable to use multi-dimensional distribution tables when possible (2005, p. 365), in order to avoid misleading conclusions sometimes possible from exclusively bi-dimensional tables. It should be noted that multi-dimensional relationships are possible within the data represented in this study. For example, the demonstrated relationship between gender and major, and major and location of previous degree may have important consequences when viewed from a three-dimensional perspective. Perhaps more importantly, variables that were not observed to be dependent in two dimensions may be revealed to have higher-order dependencies in combination with other apparently (two-dimensionally) unrelated variables. This would be an appropriate direction for further research.

Similarly, multivariate tests for the difference of means may be appropriate. As only two groups were compared over a given mean difference (i.e. elementary vs. secondary, and admissible vs. inadmissible), ANOVA techniques were not required. They are possible given the structure of the data (e.g. comparing differences in mean grades across three or more "major" groups), but such comparisons were not the focus of this thesis. Similarly, ANCOVA techniques may be successfully applied to this data (comparing groups while controlling for initial differences; Borg & Gall, 1989, p. 556). MANOVA techniques may also be instructive in identifying more complex and higher-order relationships amongst variables. Both types of analysis of variance would be appropriate for future research using this data. As this study is limited to repeated univariate techniques, it is important to note the potential for spurious findings of statistical significance. Given that the intent of the study is exploratory, and that consequences of type I error are fewer than the consequences of type II error (Stevens, 2002, p. 4-5), the results of each test of significance must be interpreted with attention to the sample sizes, the observed difference and variance (practical significance), and the probability level (statistical significance). Stevens (2002, p. 7) also notes that repeated significance tests can lead to inflated overall significance levels as illustrated in the Bonferroni Inequality:

$p_{\text{overall}} \leq p_1 + p_2 + \ldots + p_k,$ for k tests of significance

The inflated overall significance level allows for a greater likelihood of spurious findings of significant difference. While it is thus desirable to avoid performing unnecessary significance tests, significant findings with *a priori* theoretical justification and/or demonstrated practical significance can be confidently accepted (Stevens, 2002, p.

8-12). The subset of significance tests that are entirely exploratory in nature (that is, without theoretical justification) is relatively small, but it is still large enough that one or two of the findings of statistical significance will likely be spurious. Thus, the reader is again reminded to look closely at the associated statistics (sample size, practical significance, and probability) for findings of statistical significance. When statistical and practical significance are demonstrated with sufficiently large samples, one can be very confident that the observed difference in the samples reflects a real difference in the populations.

This closely relates to a common criticism of univariate exploratory studies that is sometimes called the "shotgun approach". Borg and Gall state that this is the case when measures are included in a study without any theoretical basis or common-sense rationale (1989, p. 581-582). A skeptical reader may well believe this to be the case here, as data was included on the basis of its availability. However, the primary purpose of this study is exploratory, and the inclusion of measures was made by asking the question of "Is there a reason *not* to include the measure?" rather than the converse. As such, the tested variables *could* very well be significant in illustrating differences between identified groups. Those with obvious yet unimportant differences were not tested, such as the difference in number of mathematics courses taken by elementary generalist and secondary math specialist students (as it is well understood that the secondary students are required to hold degrees in math or related fields, and would thus have a much higher average number of math courses). The study can thus be described as using a "wide-angle lens" perspective, rather than a "shotgun" approach.

Directions for Future Study

It may be interesting to determine the extent to which the sample analyzed in this study is representative of individuals in teacher preparation programs at other institutions. This could be achieved by conducting a similar study at other institutions, or sampling their populations to identify differences over what are determined to be critical parameters. It would also be instructive to extend the sample of individual students and applicants to the University of Calgary's program to include early childhood education specialists, and enrolled secondary mathematics specialists holding previous degrees from other institutions. The inclusion of additional inadmissible secondary applicants' records may give a more accurate representation of this unique population. Similarly, collection of data for inadmissible elementary applicants would enable comparison with the admitted applicants, which would be potentially informative.

Specific subsamples may be identified from within the existing data, and may be interesting to analyze. For example, the subsample of individuals who turned down offers of admission, and the subsample of students who applied to the program more than once could prove to be sufficiently unique to warrant closer examination. The existing data can also be analyzed using multivariate techniques, which could identify more subtle and/or higher-order relationships between variables.

All of the variables and samples and groups could be of tremendous further utility if data regarding subsequent teaching experience were to be collected. If, for example, a mathematics teacher effectiveness instrument were administered to individuals in the study, causal effect of the current variables could be ascertained. Subsequent interviews and focus groups with a representative sample of new teachers and/or current students could be of great service in qualifying the descriptive values and significant differences observed. Effect upon the future teaching practice could be explored, as could the possibilities for altering the teacher preparation program to directly address identified student needs. Additionally, the unintended consequences of program structural factors (e.g. the admission policy) could be identified and better understood, and addressed if necessary.

Information regarding the demographic and academic characteristics of various samples (e.g. those who declined offers of admission, those who later withdrew from the program, etc.) could prove to be useful in recruiting applicants to the program, and retaining enrolled students.

Perhaps most importantly, this study responds to the calls for research providing greater detail regarding the academic background of prospective teachers. This information can be used to further refine education production function research by considering that the academic background of prospective math teachers is highly variable from a course-by-course perspective. This detailed level of teacher content knowledge (more sophisticated than treating all math courses as equivalent) can be investigated for interaction with student achievement within particular math topics instead of "math composite scores", which has typically been the case.

Finally, other school subjects might also benefit from a detailed analysis of the coursework history of prospective teachers similar to this analysis of the coursework of prospective math teachers.

Concluding Remarks

This thesis provides a detailed analysis of the mathematics and math-related coursework history of prospective elementary generalist and secondary mathematics specialist teachers. Using coursework topics and corresponding grades as measures, it summarizes a significant portion of the mathematics content knowledge of prospective math teachers. While this responds directly to the identification of what mathematics teachers *do* know, it also helps to explore the more difficult issue of what math teachers *should* know.

Mathematics Before Teaching represents a step forward in documenting the post-secondary coursework experience of prospective mathematics teachers. It demonstrates that, as measured by coursework history, the content knowledge of individual prospective elementary and secondary teachers of mathematics is highly variable. It also shows that with regard to profiles of coursework history, there are distinct groups within the population of prospective math teachers.

This study confines its scope to the mathematics and math-related coursework of prospective math teachers. Within this context, it shows that the validity of using coursework history as a proxy measure for teacher content knowledge can be improved by increasing the level of detail in the coursework measures. It also shows that it may be appropriate to disaggregate the population of teachers in studies of mathematics teacher content knowledge in order to better represent the variation in coursework profiles of distinct groups within the population. Finally, it suggests that it may be appropriate to alter the admission criteria for the program studied to address the fact that a majority of prospective elementary generalist teachers have taken no post-secondary mathematics courses, and that the last ten courses taken by the prospective secondary math specialist teachers may not be representative of their overall coursework history.

It is hoped that this study will serve to illustrate both the importance of and the practical issues for dealing with the consideration of mathematics teacher content knowledge from the perspective of exposure to individual math topics and corresponding grades received, rather than as an aggregate whole without regard to different degrees of mastery in different content areas within mathematics. This approach allows for a much more comprehensive and interpretable use of coursework history as a measure of mathematics teacher content knowledge.

BIBLIOGRAPHY

Ahn, S., & Choi, J. (2004, April). *Teachers' subject matter knowledge as a teacher qualification: A synthesis of the quantitative literature on students' mathematics achievement.* Paper presented at the 2004 Annual Meeting of the American Educational Research Association, San Diego, CA.

Alberta Education (n.d.). *Teaching quality standard applicable to the provision of basic education in Alberta*. Retrieved November 30, 2007, from http://www.education.alberta.ca/department/policy/k-12manual/section4/teachqual.aspx

Association of Canadian Deans of Education (n.d.). *Accord on initial teacher education.* Retrieved November 30, 2007, from http://www.csse.ca/ACDE/TeacherAccord.pdf

Ball, D. L. (1990). The mathematical understandings that prospective teachers bring to teacher education. *Elementary School Journal*, *90*(4), 449-466.

Ball, D. L., Lubienski, S. T., & Mewborn, D. S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of Research on Teaching* (pp. 433-456). Washington, DC: American Educational Research Association.

Ball, D. L., Thames, M. H., & Phelps, G. (2005, April). *Articulating domains of mathematical knowledge for teaching.* Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, QC.

Bobbitt, F. (1918). *The curriculum.* Cambridge, MA: The Riverside Press.

Borg, W. R., & Gall, M. D. (1989). *Educational research: An introduction* (5th ed.). White Plains, NY: Longman Inc.

Brewer, D. J., & Goldhaber, D. D. (1996). Educational achievement and teacher qualifications: New evidence from microlevel data. In B. Cooper & S. Speakman

(Eds.), *Advances in educational productivity. Volume 6: Optimizing education resources* (pp. 243-264). Greenwich, CT: JAI Press.

Classes: What to take (2006, September). Gauntlet frosh supplement. *University of Calgary Gauntlet* (Suppl.).

Cohen, L., Manion, L, & Morrison, K. (2005). *Research methods in education* (5th ed.). New York: RoutledgeFalmer.

Crawford, J., & Impara, J. C. (2001). Critical issues, current trends, and possible futures in quantitative methods. In V. Richardson (Ed.), *Handbook of Research on Teaching* (pp. 133-173). Washington, DC: American Educational Research Association.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives, 8*(1), retrieved December 01, 2007, from http://epaa.asu.edu/epaa/vol8.html

Davis, B., & Simmt, E. (2006). Mathematics-for-teaching: An ongoing investigation of the mathematics that teachers (need to) know. *Educational Studies in Mathematics, 61*, 293-319.

Dewey, J. (1929). My pedagogic creed. *Journal of the National Education Association, 18*(9), 291-295.

Doll, W. E. (1993). *A post-modern perspective on curriculum.* New York: Teachers College Press.

Field, J. (2006, November). *Complicated conversations*. Plenary session at the International Conference on Teacher Education, Calgary, AB.

Finn, J. D., Gerber, S. B., & Wang, M. C. (2002). Course offerings, course requirements, and course taking in mathematics. *Journal of Curriculum and Supervision, 17*(4), 336-366.

Flinders, D. J., & Thornton, S. J. (Eds.). (2004). *The curriculum studies reader.* New York: Routledge Falmer.

Floden, R. E. (2001). Research on effects of teaching: a continuing model for research on teaching. In V. Richardson (Ed.), *Handbook of Research on Teaching* (pp. 3-16). Washington, DC: American Educational Research Association.

Floden, R. E., & Meniketti, M. (2005). Research on the effects of coursework in the arts and sciences and in the foundations of education. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education: The report of the AERA panel on research and teacher education* (pp. 737-759). Mahwah, NJ: Lawrence Erlbaum Associates.

Goulding, M., Hatch, G., & Rodd, M. (2004) Undergraduate mathematics experience: Its significance in secondary mathematics teacher preparation. *Journal of Mathematics Teacher Education, 6*(4), 361-393.

Grossman, P. L., Wilson, S. M., & Shulman, L. S. (1989). Teachers of substance: Subject matter knowledge for teaching. In M. C. Reynolds (Ed.), *Knowledge base for the beginning teacher* (pp. 23-36). Toronto, ON: Pergamon Press.

Hanushek, E. A., (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature, 24*(3), 1141-1177.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), pp 371-406.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal, 105*(1), 11-30.

Hopkins, K. D., Glass, G. V., & Hopkins, B. R. (1987). *Basic statistics for the behavioral sciences* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall, Inc.

Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement, 30*, 607-610.

Leinhardt, G., Putnam, R. T., Stein, M. K., & Baxter, J. (1991). Where subject knowledge matters. In J. Brophy (Ed.), *Advances in research on teaching (Volume 2): Teacher's knowledge of subject matter as it relates to their teaching practice* (pp. 87-113). Greenwich, CT: JAI Press.

Luk, H. S. (2005). The gap between secondary school and university mathematics. *International Journal of Mathematical Education in Science and Technology, 36*(2-3), 161-174.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States.* Mahwah, NJ: Erlbaum.

McDiarmid, G. W., Ball, D. L., & Anderson, C. W. (1989). Why staying one chapter ahead doesn't really work: Subject-specific pedagogy. In M. C. Reynolds (Ed.), *Knowledge base for the beginning teacher* (pp. 193-205). Toronto, ON: Pergamon Press.

McGatha, M. B., Brown, E. T., & Thompson, C. S. (2007, April). What mathematics should elementary grade mathematics teachers really know? Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review, 13*(2), 125-145.

National Board for Professional Teacher Standards (n.d.). *Five core propositions.* Retrieved November 30, 2007, from http://www.nbpts.org/the_standards/the_five_core_propositio

National Council for Accreditation of Teacher Education (n.d.). *NCATE unit standards (2006 edition).* Retrieved November 30, 2007, from http://www.ncate.org/public/unitStandardsRubrics.asp?ch=4

National Council of Teachers of Mathematics (1991). *Professional standards for teaching mathematics.* Reston, CA: NCTM.

Norusis, M. J. (2005). *SPSS 14.0: Statistical Procedures Companion.* Upper Saddle River, NJ: Prentice Hall Inc.

Pascuzzo, T. (2007, May). *The math-related academic history of prospective math teachers in a two-year after degree program.* Paper presented at the annual meeting of the Canadian Society for the Study of Education, Saskatoon, SK.

Reckase, M. D., (2007, April). *The science and sophistry of educational assessment.* Division D Vice-Presidential address given at the annual meeting of the American Educational Research Association, Chicago, IL.

Rowan, B., Correnti, R., & Miller, R. J., (2002). What large-scale, survey research tells us about teacher effects on student achievement: Insights from the Prospects study of elementary schools. *Teachers College Record, 104*(8), 1525-1567.

Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H., Wiley, D. E., Cogan, L. S., et al. (2001). *Why schools matter: A cross-national comparison of curriculum and learning.* San Francisco, CA: Jossey-Bass.

Schmidt, W. H., McKnight, C. C., Valverde, G. A., Houang, R. T., & Wiley, D. E. (1997). *Many visions, many aims: Vol. 1. A cross-national investigation of curricular intentions in school mathematics.* Norwell, MA: Kluwer Academic Publishers.

Schwab, J. J. (1978). *Science, curriculum, and liberal education.* Chicago: University of Chicago Press.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14.

Shulman, L. S. (1981). Disciplines of inquiry in education: An overview. *Educational Researcher, 10*(6), 5-12, 23.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Sullivan, P. (2003). Editorial: Incorporating knowledge of, and beliefs about, mathematics in to teacher education. *Journal of Mathematics Teacher Education, 6,* 293-296.

Tyler, R. W. (1949). *Basic principles of curriculum and instruction.* Chicago: University of Chicago Press.

Usiskin, Z. (2001, November). *Teachers' mathematics: A collection of content deserving to be a field.* Paper presented at the National Summit on the Mathematical Education of Teachers, Falls Church, VA.

Wilson, S. M., Floden, R. E., & Ferrini-Mundy, J. (2002). Teacher preparation research: An insider's view from the outside. *Journal of Teacher Education, 53*(3), 190-204.

Wilson, S. M., Floden, R. E., & Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations* (Document R-01-3). Washington, DC: University of Washington, Center For Study of Teaching and Policy.

Zeichner, K. M. (2005). A research agenda for teacher education. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education: The report of the AERA panel on research and teacher education* (pp. 737-759). Mahwah, NJ: Lawrence Erlbaum Associates.

APPENDIX A – SAMPLE RECORDS

These two fictitious exemplars show sample elementary generalist and secondary math specialist records. The record with ID# 5092 is an elementary generalist (four digits), applying for entry to the MT Program in 2005, and was the 92nd such individual recorded. The record with ID# 200705 is a secondary math specialist (six digits), applying for entry to the program in 2007, and was the 5th such individual record recorded. Both individuals took Pure Math 30 in high school, as their grades record. All coursework was taken at the University of Calgary, so only abbreviated course titles are recorded, and the grades are already on a 4-point scale. The number of mathematics and statistics and related courses taken varied from zero to 59; these samples show a condensed template. Information on the last ten courses taken by elementary students was not recorded.

ID#	5092	200705
Gender (1=male, 2=female)	2	1
Admission GPA	3.42	3.37
Degree 1 University	UofC	UofC
Degree 1 Name (type and major)	BKin	BSc AMAT
Degree 1 year graduating	2003	2005
Other Degrees	none	none
Notes	none	none
High school math grade		
Applied Math 30		
Pure Math 30	59	94
Math 31 (Calculus)		88
Coursework: names and grades 1	biol 231 C-	biol 231 C+
2	zool 361 C-	chem 201 B
3	zool 363 C	math 251 B
4	astr 205 B-	phys 221 A
5	biol 307 B-	biol 233 B+
6	glgy 209 C+	chem 203 B+
7	biol 233 A-	math 211 A
8	biol 205 A-	math 253 A
9	cpsc 225 A-	phys 223 A-
10		math 349 B+
11		math 271 A-
12		amat 311 A-
13		cpsc 215 C+
14		math 321 B
15		math 311 A

16	math 323 C+
17	stat 437 C+
18	biol 315 B-
19	amat 413 A-
20	math 411 B+
Last ten courses: 1	biol 315 B-
2	math 411 B+
3	amat 413 A-
4	stat 437 C+
5	econ 357 B
6	econ 311 A
7	econ 399 A
8	econ 303 B-
9	econ 305 A
10	econ 387 A

After codes were established for all non-numerical data, each record was transformed and recoded into the following format (neither of these individuals had more than one degree, so D2 and D3 fields are empty; code "jrl" represents a junior level life sciences course; code "sc" represents a senior level calculus course):

ID#	5092	200705
Gender (1=male, 0=female)	2	1
D1 university	1	1
D1 major	9	1
D1 level	1	1
D1 year graduating	2003	2005
D2 university		
D2 major		
D2 level		
D2 year graduating		
D3 university		
D3 major		
D3 level		
D3 year graduating		
AMAT 30		
PMAT 30	57	92
MATH 31		87
GPA - admission	3.42	3.37
Application type	0	0
Transcript status	1	1
Admission status	1	1
Coursework: names and grades 1	jrl C-	jrl C+
2	srl C-	jrp B

3	srl C-	jc B
4	jrp B-	jrp A
5	srl B-	jrl B+
6	jrp C+	jrp B+
7	jrl A-	jl A
8	jrl A-	jc A
9	jrc A-	jrp A-
10		sc B+
11		jp A-
12		sa A-
13		jrc C+
14		ss B
15		sl A
16		ss C+
17		ss C+
18		srl B-
19		sa A-
20		sl B+
Last ten courses: 1		srl B-
2		sl B+
3		sa A-
4		ss C+
5		so B
6		so A
7		so A
8		so B-
9		so A
10		so A

The coursework within individual records was collated into course categories giving the number of courses taken for each type of course. Additionally, the letter grades for coursework were converted to numerical scores, and averages (arithmetic mean) were calculated for each course category.

ID#	5092	200705
Gender (1=male, 0=female)	2	1
D1 university	1	1
D1 major	9	1
D1 level	1	1
D1 year graduating	2003	2005
D2 university		
D2 major		
D2 level		
D2 year graduating		
D3 university		

D3 major		
D3 level		
D3 year graduating		
AMAT 30		
PMAT 30	57	92
MATH 31		87
GPA - admission	3.42	3.37
Application type	0	0
Transcript status	1	1
Admission status	1	1
Number of courses: jc		2
sc		1
gc		
jl		1
sl		2
gl		
jp		1
sp		
gp		
ja		
sa		2
ga		
js		
ss		3
gs		
je		
se		
ge		
jo		
so		
go		
jrp	2	4
srp		
grp		
jrl	3	2
srl	3	1
grl		
jrc	1	1
src		
grc		
jre		
sre		
gre		
jra		
sra		
gra		
"Mathematics and Statistics" average grade		3.36
"Related" average grade	2.69	3.08
Last ten number jm		

sm	3
gm	
jr	
sr	1
gr	
jo	
so	6
go	
Last ten grade jm	
sm	3.1
gm	
jr	
sr	2.7
gr	
jo	
so	3.62
go	

APPENDIX B – ADDITIONAL TABLES

Table A1 - ***Most Advanced Degree Received by Enrolled Students***

	Frequency		Percent	
Type of Degree	Elementary	Secondary	Elementary	Secondary
Bachelor	167	69	57.8	59.5
Joint Program	17	30	5.9	25.9
Master	4	3	1.4	2.6
Doctorate	0	2	0.0	1.7
unknown	101	12	34.9	10.3
Total	289	116	100.0	100.0

Table A2 – ***High School Mathematics Grade for Successful and Unsuccessful Secondary Applicants***

	Pure Math 30		Math 31	
Standard	Successful	Unsuccessful	Successful	Unsuccessful
Mean	83.33	78.94	81.15	80.33
SD	7.302	10.207	9.64	13.747
n	36	18	26	12

Table A3 - ***Composition of Secondary Applicants' Last Ten Courses by Topic***

Number of Courses	Admissible: Math and Stats	Admissible: Related	Admissible: Other	Inadmissible: Math and Stats	Inadmissible: Related	Inadmissible: Other
None	10	30	4	8	12	1
1	5	17	9	2	5	2
2	6	2	5	1	3	1
3	15	5	9	5	0	6
4	9	1	8	2	3	6
5	9	2	4	3	2	2
6	6	1	16	4	0	2
7	5	1	6	2	2	2
8	0	4	3	1	0	3
9	3	4	2	0	0	1
10	0	1	2	0	1	2

Table A4 - ***Composition of Secondary Applicants' Last Ten Courses by Level***

Number of Courses	Admissible: Junior	Admissible: Senior/Graduate	Inadmissible: Junior	Inadmissible: Senior/Graduate
None	14	0	8	0
1	18	0	8	0
2	18	0	3	0
3	8	1	5	0
4	5	1	2	1
5	3	3	1	1
6	1	5	1	2
7	1	8	0	5
8	0	18	0	3
9	0	18	0	8
10	0	14	0	8

APPENDIX C – ETHICS APPROVAL

UNIVERSITY OF
CALGARY

CERTIFICATION OF INSTITUTIONAL ETHICS REVIEW

This is to certify that the Conjoint Faculties Research Ethics Board at the University of Calgary has examined the following research proposal and found the proposed research involving human subjects to be in accordance with University of Calgary Guidelines and the Tri-Council Policy Statement on *"Ethical Conduct in Research Using Human Subjects"*. This form and accompanying letter constitute the Certification of Institutional Ethics Review.

File no:	**5197**
Applicant(s):	**Tony M. Pascuzzo**
Department:	**Graduate Division of Educational Research**
Project Title:	**The Mathematics Academic Background of Prospective Teachers in a Two-Year After-Degree Program**
Sponsor (if applicable):	

Restrictions:

This Certification is subject to the following conditions:

1. Approval is granted only for the project and purposes described in the application.
2. Any modifications to the authorized protocol must be submitted to the Chair, Conjoint Faculties Research Ethics Board for approval.
3. A progress report must be submitted 12 months from the date of this Certification, and should provide the expected completion date for the project.
4. Written notification must be sent to the Board when the project is complete or terminated.

Janice Dickin

26 April 2007

Janice Dickin, Ph.D, LLB,
Chair
Conjoint Faculties Research Ethics Board

Date:

Distribution: (1) Applicant, (2) Supervisor (if applicable), (3) Chair, Department/Faculty Research Ethics Committee, (4) Sponsor, (5) Conjoint Faculties Research Ethics Board (6) Research Services.

2500 University Drive N.W., Calgary, Alberta, Canada T2N 1N4 • www.ucalgary.ca