4-1-2017

Teacher Characteristics and Student Achievement in Mathematics in International Schools

Michael Arcidiacono
marcidiacono04@georgefox.edu

This research is a product of the Doctor of Education (EdD) program at George Fox University. Find out more about the program.

Recommended Citation
http://digitalcommons.georgefox.edu/edd/96

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Digital Commons @ George Fox University. It has been accepted for inclusion in Doctor of Education (EdD) by an authorized administrator of Digital Commons @ George Fox University. For more information, please contact arolfeg@georgefox.edu.
TEACHER CHARACTERISTICS AND STUDENT ACHIEVEMENT IN MATHEMATICS IN INTERNATIONAL SCHOOLS

by

Michael Arcidiacono

FACULTY RESEARCH COMMITTEE

Chair: Patrick Allen, PhD
Member: Dane Joseph, PhD
Member: Susanna Thornhill, PhD

Presented to the Faculty of the
Doctor of Educational Leadership Department
George Fox University
in partial fulfillment for the degree of

DOCTOR OF EDUCATION

APRIL 13th, 2017
"TEACHER CHARACTERISTICS AND STUDENT ACHIEVEMENT IN MATHEMATICS IN INTERNATIONAL SCHOOLS," a Doctoral research project prepared by MICHAEL ARCIDIACONO in partial fulfillment of the requirements for the Doctor of Education degree in Educational Leadership.

This dissertation has been approved and accepted by:

4/13/17  Patrick Allen, PhD
Date      Committee Chair

4/13/17  Dane Joseph, PhD
Date      Professor of Education

4/13/17  Susanna Thornhill, PhD
Date      Associate Professor of Education
Abstract

This study explored the relationship between traditional measures of teacher status (experience, level of education and certification) and student achievement growth in mathematics in international schools in Brazil. These measures have been explored extensively in educational research in the United States. However, similar studies have not been conducted in international school settings. The literature review investigated the existing body of literature on teacher status as well as the nature of mathematics achievement in international schools and the rise in use of value-added models as a means of measuring growth. Multiple regression analysis was used to determine the impact of teacher professional characteristics on student achievement growth in mathematics. A significant challenge that was identified and discussed in the study was the difficulty in collecting data and conducting quantitative research in international schools. This challenge limited the findings of the study; nevertheless, some findings from the study intend to initially deepen an understanding of some of the unique characteristics of international schools and provide a basis for future research. The results of the study therefore added to the current body of research on the relationship between teacher status and student achievement growth and also added to the small body of research on international schools.

Keywords: International schools, mathematics achievement, teacher status
Acknowledgements

The road to completing a dissertation is long, winding and bumpy. Just to make things more interesting, I undertook this journey while my wife and I were in the midst of raising our very young children. I am very grateful to my wife Julie for supporting me through this process and doing so much in dedication to our family in what must have seemed like an interminable process at various points along the way.

I depend on God and my family for continual support, and am fortunate to never be let down. From the example set by my late father (the first in the family to write a dissertation), to the love and support always provided by my mother and brothers, I know that I am truly blessed. I am particularly grateful as well to my brother Peter who helped me on many occasions to think through my research, and served as a sounding board at the stage of analysis to help keep me on track.

I would like to thank my committee chair Dr Patrick Allen for his continual positive support – it was needed during the bumpy parts of the road, of which there were many! I always appreciated the words of encouragement that were in each email. I am also especially grateful to Dr Dane Joseph, for helping me to understand better how to design quantitative research and how to conduct the analysis. It took me quite a while to grasp how to do certain elements and Dr Joseph was patient and supportive throughout. Finally, I would like to thank Dr Susanna Thornhill for her feedback and guidance in helping me to organize my thinking more clearly on the literature review.
# Table of Contents

ABSTRACT .......................................................................................................................... ii  
ACKNOWLEDGEMENTS ................................................................................................. iii  
TABLE OF CONTENTS ..................................................................................................... iv  
CHAPTER 1 ......................................................................................................................... 1  
   INTRODUCTION ............................................................................................................... 1  
      Background to the Problem ....................................................................................... 1  
      Statement of Purpose ................................................................................................. 3  
      Significance of the Study ........................................................................................... 4  
      Research Questions ................................................................................................... 4  
      Key Terms .................................................................................................................. 5  
      Limitations and Delimitations .................................................................................... 6  
CHAPTER 2 ......................................................................................................................... 9  
   REVIEW OF THE LITERATURE .................................................................................... 9  
      Introduction ............................................................................................................... 9  
      Years of experience and student achievement in mathematics ......................... 11  
      Level of degree and student achievement in mathematics ................................ 12  
      Certification and student achievement in mathematics ........................................ 14  
      Measuring achievement in mathematics .................................................................. 15  
      The nature of MAP testing ....................................................................................... 20  
      Value-added models ................................................................................................... 22  
CHAPTER 3 ......................................................................................................................... 28  
   METHODS ....................................................................................................................... 28  
      Research Questions ................................................................................................... 28  
      Settings & participants ............................................................................................... 28
TEACHER CHARACTERISTICS IN INTERNATIONAL SCHOOLS

Administration & data collection .................................................. 30
Determining eligible teachers and their classes ............................. 31
Collecting the class data .............................................................. 31
Collecting the teacher data .......................................................... 32
Analytical procedures ............................................................... 32
Variables ................................................................................... 33
Research Ethics .......................................................................... 36

CHAPTER 4 .................................................................................. 38
RESULTS .................................................................................... 38
Descriptive statistics of participants .............................................. 38
Testing assumptions of multiple regression ................................. 43
Multiple regression analysis ....................................................... 44

CHAPTER 5 .................................................................................. 49
DISCUSSION & CONCLUSIONS .................................................. 49
Research questions – discussion & results ..................................... 50
Limitations of the study ............................................................... 53
Implications for future research .................................................. 54

References ................................................................................. 56
Appendices .................................................................................. 62
Appendix A: Sample data collection spreadsheet .......................... 63
Appendix B: Sample ASG class report .......................................... 64
Appendix C: General teacher survey ............................................ 65
Appendix D: Teacher demographics (figures) ................................. 67
Figure 1: Box plot (bachelor’s degree) .......................................... 67
Figure 2: Box plot (master’s degree) ............................................ 67
Figure 3: Box plot (country of certification)............................... 68
Appendix E: Assumptions of multiple regression.......................... 69
Figure 4: Partial regression plot (experience) ............................. 69
Figure 5: Partial regression plot (tenure) ................................. 69
Figure 6: Studentized residual plot........................................ 70
Figure 7: P-P plot .............................................................. 71
Chapter 1

Background to the problem

International schools have experienced a period of tremendous growth in the past 20 years. Brummitt (2009) noted a change in the number of schools from 2584 in the year 2000 to over 5187 schools in 2009. Part of this increase was a result of a widening in the group of students international schools aim to serve. In attempting to define the concept of an international school, Hayden (2011) noted a historical understanding of a not-for-profit school devoted to serving the children of expatriates. However, she wrote that the concept of an international school has grown to encompass for-profit schools as well as schools that target the upper or aspiring middle class of the host country.

Regardless of the reasons for such expansion, the growth rates have been staggering, especially for school administrators who must compete to identify and hire teachers to keep up with the growth. According to the International School Consultancy (ISC), the number of required teachers for international schools worldwide will jump from 300,000 in 2013 to 529,000 by 2022. And while it is one thing to identify and hire qualified teachers, as difficult and expensive as that process may be, it is certainly another thing altogether to retain them.

In fact, international school teacher turnover and retention rates are similar to schools in the United States. Odland & Ruzicka (2009) cited a joint survey of the Council of International Schools (CIS) and the European Council of International Schools (ECIS) who found teacher turnover amongst 270 member schools to be 14.4% for the 2005/6 school year. A survey of international schools in the Near East South Asia (NESA) region found teacher turnover for the period 2006 – 2009 to be even higher, averaging 17% (Mancuso, Roberts, & White, 2010). However, this study reached a much smaller population of schools, with roughly 10% of the participating schools than the CIS/ECIS survey. The School and Staffing Survey (SASS), conducted by the National Center for Education Statistics in the United
States, found the turnover amongst public teachers to average between 15% and 17% in their three most recent surveys since 2004.

Even though the retention and turnover rates in international schools are roughly comparable to US public education, it is a major concern to international school administrators given the resources required to recruit and relocate new teachers. Skinner (1998) described various financial costs involved in hiring educators to teach abroad—flying to hiring fairs, hotel costs, providing housing, etc. Another option exists for some international schools, to employ locally contracted teachers who are typically nationals of the country where the school is located. Odland & Ruzicka (2009) noted that locally contracted teachers are usually paid less than teachers from abroad, and have fewer benefits.

Having qualified and effective teachers on staff in international schools is naturally of great importance. When considering the significance of the teacher, Marzano (2003) wrote, “We live in an era when research tells us that the teacher is probably the single most important factor affecting student achievement—at least the single most important factor that we can do much about” (p. 1). A significant body of research has investigated the relationship between teacher characteristics and student achievement in United States settings (Brewer & Goldhaber, 1997; Brewer & Goldhaber, 2000; Buddin & Zamarro, 2009; Greenwald, Hedges, & Laine, 1996; Hanushek, 1997; Heck, 2007; Huang & Moon, 2008; Lefgren & Sims, 2012; Rivkin, Hanushek, & Kain, 2005; Rowan, Chiang, & Miller, 1997; Rowan, Correnti, & Miller 2002). However, research of this nature has not yet been conducted in international schools.

One of the challenges of conducting quantitative research in international schools is the collection of data. For context, there are 14 schools in Brazil registered with the Association of American Schools in South America (AASSA). This large organizing body for the region promotes conferences for students, offers training for teachers and
administrators, and generally creates other opportunities for schools to share practices. And all 14 of these schools also hold accreditation status with AdvancED or the New England Association of Schools and Colleges (NEASC), two accreditation bodies who allow international schools to offer an American diploma to their students. However, none of these governing bodies or organizations keep or pool student achievement data.

Fortunately, there is a common assessment tool that is used by many international schools in Brazil. The assessment tool is the Measures of Academic Progress (MAP) test from the Northwest Evaluation Association (NWEA). Of the 14 schools mentioned above, 12 offered the MAP testing to their students during the 2015-16 school year. Therefore, an opportunity existed to extend the current body of research and investigate how teacher characteristics (e.g., years of experience, level of degree, etc.) relate to student achievement in international schools.

Statement of purpose

The purpose of this study was to examine relationships between teacher characteristics and student achievement in international school settings. A survey instrument was developed to gather information on teacher’s professional characteristics (years of teaching experience, level of education, etc). Achievement in mathematics was defined as the growth a student made between the Fall 2015 MAP testing session and the Spring 2016 testing session.

The study was regional in nature, focusing on international schools in Brazil who used a common mathematics achievement test during the 2015-2016 academic calendar year. Several variables related to a teacher’s professional characteristics were operationalized for the study. Multiple regression analysis was utilized to determine the impact of each of the independent variables on student mathematics achievement.
For the purpose of this study, student achievement was narrowly focused on achievement in mathematics (the MAP assessment tests understanding in mathematics and English). This was for two reasons. First, the educational background of the researcher was in mathematics, so there was an element of personal interest in conducting research in this area. Secondly, the role of language in international schools may impact different subjects in different ways. Through the experience of working and teaching in Brazil for 10 years, the researcher found that the international schools had very high levels of “host-country nationals”—Brazilian students who followed an international curriculum in English. As these students were by nature second language learners, restricting the study to student achievement in mathematics addressed some of the variation in students’ levels of fluency in English. That is, in some respects, mathematics was a second language for all students, which effectively equalized student achievement.

Significance of the study

This study was significant for two reasons. First, it broadened the research that existed on the relationship between teacher characteristics and student achievement by including data from a context previously unexplored—the international school context. The existing body of research of this relationship was significant, yet it was a result of studies based on data collected primarily in the United States. Second, this study added to the currently small base of research that had been conducted in international schools. As the number of international schools continues to grow, more research is needed in this context.

Research Questions

1. Does the overall number of years of teacher experience have an impact on student achievement in mathematics?

2. Does the number of years of teaching experience at the current school have an impact on student achievement in mathematics?
3. Does the location of the teacher’s certification have an impact on student achievement in mathematics?

4. Does the teacher’s level of education have an impact on student achievement in mathematics?

**Key Terms**

*Expatriate teacher* -- An overseas contract teacher is someone who, at the point of employment, has been residing in another country and will need to physically move to the new place of employment in the country where the international school resides. A teacher of this profile will have naturally completed their teacher training in their home country or somewhere else abroad.

*Gains model* -- A model of measuring achievement in students by looking at the gain made from one assessment to the next. This model is typically employed in consecutive academic years to measure a student’s growth from year to year, or in the same academic year to measure growth from the beginning of the year to the end of the year.

*Host country national* – A citizen of a country where an international school is located. In the context of an international school, the citizen may be a student attending the school, or a teacher at the school.

*International school* – Historically, the concept of an international school has been difficult to define. For the purpose of this paper, the definition is based on the following four points: may serve a varied local and expatriate community, may attract students world-wide, may be privately owned or governed by a board that includes parents, and may pay fees or use scholarships for students (Hill, 1994, as seen in Hayden & Thompson, 1995).

*Local expatriate teacher* -- A local expatriate teacher is defined as someone who is typically from an English speaking country, and who has elected to make the country where
the international school resides as their home. The teacher training of a local expatriate typically has, in virtually all instances, been completed in their home country.

*Local contract teacher* -- A local contract teacher is defined as a host country national of the country where the international school resides. In most cases, their teacher training has also been completed in the host country.

*Multicollinearity* - A phenomenon in which two or more predictor variables in a multiple regression model are highly correlated, meaning that one can be linearly predicted from the others with a substantial degree of accuracy.

*National school or system* – For the purpose of this paper, a national school or system of schools is considered located in the United States.

*Norm-referenced test* -- A test that measures the knowledge of an individual of a particular subject, but reports this measurement within the context of the overall testing population’s knowledge of that subject.

*Value Added Model (VAM)* -- A method of assessing teacher effect on student achievement by considering the difference in achievement from one test to another. This may occur from one year to the next, or in the same year but in a pre-test/post-test manner.

*Reliability* -- The measure of how consistent results of a test would be, if the test were administered several times under similar conditions.

*Validity* -- The measure of how well a test measures what it is intended to measure.

**Limitations and Delimitations**

A limitation of this study was the inability to control for school level factors, such as how students were assigned to classrooms. An assumption was made that children were randomly assigned to mathematics classes and were not grouped in classes by ability. However, these assumptions could not be verified. For this reason, Rivkin, Hanushek, and Kain, (2005) note that making between class comparisons can be problematic.
Another limitation was that the design of this study did not account for differences in the socio-economic background of the students involved. Student socio-economic data was not publicly available and thus was not collected for this study. McCaffrey et al. (2003) wrote that not accounting for variation amongst the student population could skew the effect of the teachers being investigated.

Finally, two limitations existed that affected how the results of this study could be generalized to the wider universe of international schools. The first was that the study was regional in nature, focusing on the population of international schools in Brazil. International schools outside of Brazil may have populations of teachers from different backgrounds or use assessments other than the MAP test. Secondly, random probability sampling of teachers and their classes was not conducted. Rather, only teachers (and their classes) who were willing to participate and fill out the survey were included in the study. Therefore, this may have resulted in participation bias.

The restriction of collecting the data from the current academic calendar year resulted in a delimitation in the choice of mathematical model that was used in measuring student academic achievement. As the student achievement data was collected from Fall to Spring testing sessions of the current calendar year, a “gains model” was employed. McCaffrey, Lockwood, Koretz, & Hamilton (2003) wrote that an assumption of the gains model is that the effects of previous teachers do not affect future gains. However, Rivkin et al. (2005) noted a limitation of the gains model in that the “aggregate effects” of teachers a student had in previous academic years may not be adequately accounted for in the pre-test data point, and therefore may impact the post-test result.

The restriction of collecting the data from grades 3-8 was a second delimitation of the study. In personal communication with administrators from the participating schools, grade levels below Grade 3 were not considered because the interference of language
development—it was felt that this interference would particularly more impactful in these grade levels. As most participating students were learning in a second language, their performance on a mathematics achievement test may have been hindered as a result of their ability in the English language. Above Grade 8, some students may have already earned achievement results at the top of the vertical scale. Schools tended to test fewer students at these upper grades, therefore the possible sample size for the study was further reduced.

A final delimitation existed in the chosen method of collecting the data. An administrator from each participating school was asked to assist in administering the survey and collecting the student achievement results. The involvement of an additional person(s) increased the chance of human error, or misinterpretation of the content or intent of the teacher survey by the end users (the teachers). The researcher attempted to mitigate this delimitation by working in close contact with the administrator of each school to ease the burden of data collection and reduce error in the process.
Chapter 2

Review of the literature

Notions of increased pay based on advanced degrees, years of teaching experience and appropriate certification have historically led policymakers to assume that relationships exist between these traditional teacher characteristics and the quality of the teacher themselves. These notions were brought to the fore and made explicit during the No Child Left Behind (NCLB) legislation, and the advent of Highly Qualified Teacher (HQT) status. According to the US Department of Education (2004), “To be deemed highly qualified, teachers must have: 1) a bachelor's degree, 2) full state certification or licensure, and 3) prove that they know each subject they teach (Terms to know section, para. 1). Hence, a significant body of literature exists on these teacher characteristics.

However, these assumed relationships have found mixed and, in some cases, no support in the literature on improving student achievement. For example, several studies (Buddin & Zamarro, 2009; Huang & Moon, 2008; Rivkin et al., 2005) and meta-analyses (Greenwald et al., 1996, Hanushek, 1997) have noted the lack of a consistent relationship between a teacher’s level of education and student achievement. Hanushek’s work in this area has been particularly significant. He conducted successive meta-analyses in 1981, 1986, 1989 and 1997, building up to a data set of almost 400 studies that were investigated in the 1997 publication. From these studies, he consistently found no evidence that teachers with higher degrees were more effective (in terms of student achievement) than their colleagues.

Additionally, the nature of how student achievement is measured has changed over time. The shift to measuring achievement in terms of growth, known as value-added modelling (VAM), has prompted the development of new assessments like the MAP test and Smarter Balanced that measure student achievement on a single scale. Hanushek (1997) and Rowan et al. (2002) noted the superiority of using VAMs to isolate the teacher effect on
student achievement. However, several studies have discussed limitations in VAMs (Buddin & Zamarro, 2009; Kersting, Chen, & Stigler, 2013; McCaffrey et al., 2003; Murphy, 2012; Rivkin et al., 2005; Schochet & Chiang, 2013). As well, there is mixed opinion on what steps related to school policy can or should be taken based on the results of VAMs (Kane & Steiger, 2014; Kersting, Chen, & Stigler, 2013; Koedel, Mihaly, & Rockoff, 2015).

For example, Hanushek (1997) identified from a meta-analysis of a set of studies that there was a positive (though small) relationship between a teacher’s years of experience and student achievement. When he separated the studies that employed value-added techniques from those that did not, he found that the support for this relationship increased. Other studies have also found support for the relationship between teacher experience and student achievement, (Buddin & Zamarro, 2009; Greenwald et al., 1996).

Greenwald et al. (1996) conducted a meta-analysis of studies and examined the professional characteristics of teacher education and experience. They found support for both characteristics having a significantly positive relationship with student achievement. However, Hanushek’s (1997) meta-analysis contradicted this finding for teachers’ level of education. Again, he separated the studies that used value-added techniques but still did not find support for this relationship. Buddin and Zamarro (2009) looked extensively at the test results of all three tests that are mandatory for teacher certification in a sample of elementary teachers in the Los Angeles public school district. They found no correlation with student achievement.

The mixed results evident in studies on teacher characteristics set a context for a deeper and more specific look into how or what relationships exist between the characteristics of teachers and the mathematics achievement of their students. This literature review will therefore explore several aspects of this presumed relationship. The first aspect of this literature review is an investigation into studies that have researched how these
traditionally understood proxies of teacher quality have been related to student achievement in mathematics. The nature of how achievement in mathematics is measured, particularly in international schools, is also explored. Finally, this literature review examines at the nature and methodology of both the MAP test (the assessment to be used in this study) and value-added models more generally.

Years of experience and student achievement in mathematics

Several studies have investigated the relationship between teacher characteristics such as years of experience and student achievement amongst math teachers. Rivkin et al. (2005) found that experience was important amongst early-career math teachers, noting that experienced teachers performed much better than teachers in their first year, as well as compared to second or third year teachers (however, to a slightly lesser extent). However, past these first years of teaching, they found no differences amongst the performance of teachers with more than three years of experience.

There are several strengths and limitations to the study conducted by Rivkin et al. (2005). The researchers utilized a robust data set and were able to control for the effects of variation amongst student background and also variation amongst schools. They used a value-added model and investigated a data set that spanned several years, tracking the results of students over time. However, they were not able to match individual teachers to specific students, instead calculating averages for teachers professional characteristics in each grade level for each school.

Klecker (2002) investigated data taken from a national assessment in the United States to examine the relationship between mathematics teachers’ years of experience and student achievement in 4th and 8th grade. Klecker’s study grouped teachers based on years of experience into five separate groups and then looked for significant differences in the performance of students between the groups. Klecker found a small positive relationship
between teacher experience and student achievement. However, this study did not however measure actual growth in the achievement of students. Nor were there any controls for variation amongst the population of students or the characteristics of schools.

Greenberg, Rhodes, Ye, and Stancavage (2004) drew from the same national data set for 8th graders as Klecker (2002) but addressed school level and student level factors in their methodology. They split years of experience into only two categories, whether or not a mathematics teacher had at least five years of teaching experience. Their initial findings yielded support for this relationship; however, when they controlled for student level factors the results were no longer statistically significant. Again, the limitation of a lack of longitudinal growth measure was of note in this study.

Rowan et al. (2002) used a large, multi-state data set to replicate and extend research to look at certain characteristics of elementary teachers and their relationship with student achievement. In the process of their analysis, the researchers commented on the evolution of research from measures of achievement status (at a fixed point in time) to several different variations of value-added models, noting the limitations in each model. They then utilized a model that looked at several years of achievement data for each student while controlling for school level factors, student level factors, and variation across the years of achievement. From their model, they found a positive relationship between teacher experience and student achievement in mathematics.

**Level of degree and student achievement in mathematics**

Brewer and Goldhaber (1997) studied the effect of a bachelor’s degree or master’s degree in mathematics on the achievement results of 10th graders in mathematics. They used a national data set that measured achievement status (single data point in time), though the data set did allow for the researchers to control for student level factors. When Brewer and Goldhaber stipulated that the certificate or degree was in mathematics, a significantly positive
relationship was found between this teacher characteristic and student achievement. However, when they did not specify the degree, no significant relationship was found in the data. Rowan, Chiang, and Miller (1997) utilized the same data set for their research. In their study, they factored in students’ prior achievement at the 8th grade version of the assessment as an additional student level factor. Their findings corroborated those of Brewer and Goldhaber.

Greenberg et al. (2004) found similar positive results when mathematics teachers held at least a major or minor degree in mathematics. In a recent, single-state study using longitudinal data (a value-added approach), Shuls and Trivitt (2015) also found positive results for teachers with stronger educational backgrounds in mathematics. Their study, like many others that have been cited in this section, focused on student achievement in the high school years.

In an extension of their previous research, Brewer & Goldhaber (2000) found a positive relationship as well with the amount of coursework undertaken in a teacher’s mathematics degree and the achievement of their students. The positive results of these studies were restricted to the achievement of middle or high school age students; Eberts & Stone (1984) did not find results that supported the above conclusions when examining the achievement of 4th graders.

Interestingly, Rowan et al. (2002) found a negative relationship for teachers with an advanced mathematics degree and the achievement of their students when compared with teachers who only had a bachelor’s degree. Their study was conducted using elementary school data; they hypothesized that advanced degrees could actually have a negative effect and therefore be unnecessary at that level of schooling. As the study in this dissertation incorporated data from both elementary and secondary teachers, this finding by Rowan et al. (2002) would be examined again.
Certification and student achievement in mathematics

Greenberg et al. (2004) also investigated the relationship between teacher certification and student achievement in their study. They found that students achieved higher results when they were taught by mathematics teachers who were certified. Lubienski, Lubienski, and Crane (2008) used data from the same national test as Greenberg et al. but from a more recent and expansive version. They also found a positive relationship between teachers who were certified in mathematics and the achievement of their students in 4th grade and 8th grade.

Shuls and Trivitt (2015) found contrasting findings in their study when they compared the results of teachers who were certified in a traditional manner versus teachers who went through alternative certification programs. They limited their study to early-career teachers, who had been teaching for five years or less. Their findings indicated no significant differences in the student achievement of either teaching group. However, their study did not specify any additional subject-specific certification (such as in mathematics), but only looked at traditional versus alternative pathways.

Rowan et al. (2002) investigated whether teachers who had a special certification in the subject area of mathematics had an impact on student achievement. They found no significant relationship, although they noted that teachers who held this type of certificate constituted a very small percentage (6%) of the overall teaching population in their study. As the Greenberg et al. (2004) study was done using 8th grade data and the Rowan et al. (2002) study used elementary grade level data, these findings lend support to the notion that specialized certification may be more impactful for teachers who are working with older students.

The importance of teachers and their impact on student achievement has thus been deeply researched from many different perspectives. More recent studies have benefitted...
from improved methodology, particularly in the use of value-added models, as researchers have attempted to identify what characteristics of teachers have an impact on student achievement. The studies that have focused more specifically on mathematics teachers have yielded more results, which informs the purpose of this investigation. The nature of how achievement in mathematics is measured must therefore be explored.

**Measuring achievement in mathematics**

Mathematics achievement can be measured in several ways and is dependent on the purpose for which the measures will be used. Olson (1999) gave two general purposes: one is diagnostic (or formative)—to be able to improve future instruction or learning; the second purpose refers to a question of accountability—to assess what students know so that teachers (and students) are held responsible. Regardless of the purpose, assessments must be accurate in measuring what they set out to measure (validity) and do so consistently over time (reliability). Implicit in the second form of assessment are comparisons that can be made about how a student has performed in relation to a larger group of students, known as “norm-referenced” comparisons.

Many forms of math achievement tests therefore set out to assess the content students are expected to know at a particular year level. State assessments are good examples of this, as well as assessments that are used in international settings. The Trends in International Mathematics and Science Study (TIMSS) tests student understanding in 4th and 8th grade, to make comparisons of student achievement across both national and international contexts, and to assess growth trends. Rindermann and Baumeister (2015) investigated the relationship between TIMSS questions and the content it purported to measure by using external raters drawn from teachers of the subjects measured by TIMSS as well as university psychology students. The raters found support for the relationship between the test and the curriculum, demonstrating high levels of inter-rater reliability.
Cambridge International Examinations (CIE) provide exams in many international schools around the world. According to their website, over 10,000 schools in more than 160 countries administer their tests. One of their exams, the International General Certificate of Secondary Education (IGCSE), measures student understanding of a syllabus that typically spans two years of study over grades 9 and 10, and is the traditional benchmark for students to finish their required schooling in the United Kingdom. Jones, Swan & Pollitt (2013) criticized the IGCSE test as measuring achievement primarily through questions of low levels of complexity. Their investigation applied a method for testing problem solving skills and acknowledged that the application of this method was inappropriate for a test that is intended to show understanding of content.

The SAT and ACT measure mathematics achievement differently from the previous examples. The purpose of these tests is primarily predictive, to assess college readiness. The content assessed is not taken directly from a specific high school curriculum; rather, the College Board website stresses that they rigorously conduct a process of curriculum alignment to various state/national standards when they test for content validity. According to a 2014 College Board report on the test characteristics of the SAT, the reliability coefficient was between 0.92 and 0.94, which was considered to be a high measure of reliability (Gay, Mills, & Airasian, 2012). As well, a summary report issued by the College Board cited several studies on the predictive validity of the SAT as a measure of future academic success, noting relatively strong correlations with college GPA and graduate school examinations (Shaw, 2015).

The SAT, ACT, and TIMSS all make use of multiple choice questions in their assessment of mathematical understanding; multiple-choice questions have also been the hallmark of standardized testing and are a source of much debate (Kobrin, Kim & Sackett, 2012). These researchers investigated the effects of the multiple choice questions on the SAT
versus the student-produced response questions on different measures of college readiness. They found that when the level of difficulty of the question(s) was controlled for (the student-produced response questions had higher levels of complexity, by design), the effects of the two types of question responses were very similar. In contrast, Stanger-Hall (2012) investigated the effects of test format and found that students who experienced a mixed format incorporating multiple choice questions and constructed response questions had higher levels of achievement on questions involving higher level thinking skills than students who experienced multiple choice questions only.

Several organizations who provide assessments for international schools measure mathematics achievement using constructed response or models that incorporate different approaches to assessment. The IGCSE exams comprise a mixture of short answer and longer problems with multiple parts to test student understanding. The TIMSS mathematics exam incorporates short or extended response questions in one-third of their assessment. The International Baccalaureate (IB) uses criterion related assessment in all subjects, including mathematics. They use a combination of internal assessment (performance tasks, investigations and projects) and external assessment (exams of short and extended response questions) to come up with a balanced assessment of student understanding. Since there is a norming feature to the setting of the levels of criteria, the assessment model is not purely criterion-referenced.

Regardless of the purpose or the nature of the test questions, all the examples of mathematics achievement testing given thus far are standardized, meaning they are the same test given under the same conditions to all test-takers. Several studies have been conducted and articles written on the topic of the validity of standardized assessments in mathematics (Herman, Webb, & Zuniga, 2007; Porter, 2002; Webb, Herman, & Webb, 2007) and of standardized assessments more generally (Webb, 2007). Checking the validity of the
assessment implies that the underlying relationship between the standards being tested is assessed against several criteria.

In the case of Webb (2007), there were four criteria that must be considered. The first was categorical concurrence, the extent to which there is alignment between the general categories in the assessment and the categories into which the standards are organized. The second was consistency of depth of knowledge (DOK), which is the extent to which the level of depth of knowledge in the assessment corresponds to the level of depth of knowledge required in the content strand(s). The third was the range of knowledge correspondence. Exactly as the name implies, it is the extent to which the overall breadth of knowledge across the standards is accurately represented in the assessment. The final criteria was balance of representation, to check that not only is there a breadth of knowledge in both standards and assessment but also that the balance within the content areas of each are in line with each other.

Recent trends have seen a more comprehensive approach to mathematics assessment. The Smarter Balanced Assessment Consortium (SBAC) (2015) has designed assessments that achieve both the purposes outlined by Olson (1999) and also serve as predictive measures of college readiness. Their technical report provided a comprehensive look at the design of their assessments. This included how they address content alignment and content validity, drawing on Webb’s (2007) DOK framework, and how their response format was designed - a mixed approach involving a computer-adaptive test (CAT) and a performance task. Within the CAT, multiple choice and constructed response were solicited in formats seen in Stanger-Hall (2012) and in the SAT (Kobrin, Kim, & Sackett, 2012). For the performance task, raters were used for scoring and mechanisms were in placed to check for inter-rater reliability in a manner similar to those seen in Rindermann and Baumeister (2015).
Bennett (2015) noted the advent of computer adaptive testing (CAT) as an evolutionary step that moved beyond the accountability focus of traditional standardized measures of student achievement. A joint brief issued by Pearson, ETS and the College Board (ETS, 2010) outlined some of the positive and negative aspects of this method of testing. The great benefit of CAT was the ability to track and measure growth over time, whereas traditional testing methods measured achievement at a single point. An additional strength of the adaptive model cited in the brief and noted by Bennett (2015) and SBAC (2015) was the potential for students who normally perform at the extremes (either at a very high or low level) to access a test that more accurately allowed them to demonstrate their understanding because the test accessed content from multiple year levels on its vertical scale.

The brief also outlined several negative aspects of computer adaptive testing. While perhaps a shorter test length was desirable, they noted that poor performance early on in an adaptive test may have affected a student’s ability to access more challenging questions that provide a more accurate picture of their understanding, whereas traditional testing measures are not impacted by the nature of this test design. They cite no research in support of this criticism. Rather, they simply stated the need for research to be conducted in this area. An additional concern they raised is the ability of adaptive tests to assess student understanding in non-dichotomous questions, again citing the need for further research. Finally, they viewed the greatest challenge of CAT is the construction of a test design that allows for accurate interpretation of results across test-takers, a key feature of traditional testing (ETS, 2010).

In summary, measuring achievement in mathematics has evolved significantly over time, to serve several purposes. The manner in which mathematics achievement is assessed has also changed, and the test design, particularly in the case of CAT, allows for the objectives of formative, summative, and predictive analysis to happen concurrently.
International schools are also moving in this direction as they incorporate CAT assessments such as the Measure of Academic Progress (MAP) test, created by the NWEA.

**The nature of MAP testing**

According to the NWEA website (N.D.), The MAP test was first developed for use in 2000. The MAP test assesses student achievement using the Rasch model, where students earn a RIT (Rasch Unit) on a single vertical scale that is constructed of equal intervals. The RIT score is determined by measuring the level of difficulty of a question such that a student should be able to answer it roughly 50% of the time (NWEA, 2015). McCaffrey et al. (2003) noted the importance of the use of a vertical scale to be able to draw inferences about teacher effects on student achievement regardless of the grade level.

The NWEA website (N.D.) noted that the content of the MAP test is aligned to state standards, balanced across specific tested areas. Porter (2002) and Webb (1997) wrote that alignment is a critical feature connecting assessment to state standards. The MAP test incorporates content from year levels above and below the tested year level to be able to test a student’s true achievement level by using adapted-response questions, based on whether or not the student had answered the previous question(s) correctly. Hence, the MAP test is not a standardized test as each student does not face the exact same test. Rather, the MAP is a CAT; each test is created from a common test bank and is unique to each student.

The NWEA website (N.D.) also details how they address construct validity in the MAP test. They do so at three levels: at the design of the test itself, at the level of the student experience, and at the question level. Construct validity at the design level relates to what is included in the test (the content standards of the current year, and the years preceding or following) and the balance of cognitive complexity (using the DOK model). The MAP test makes use of the Webb (2007) model to check for content alignment. At the level of the student experience, the NWEA (N.D.) notes that the difficulty of questions presented, based
on student performance, is appropriate to the student’s zone of proximal development. Finally, construct validity at the question level is assessed to check that the cognitive complexity of the question itself matches the complexity required in the content standard.

The NWEA (2004) has published reports treating the questions of reliability and content validity of the MAP test. They showed the validity of the test to have a correlation typically ranging between 0.80 to 0.95 across almost all year levels and sub-sections of the test. This report also provided evidence of similarly strong correlation coefficients for the content validity of the test in comparison to more than ten state standardized assessments. In a multi-year, multi-state study, Wang, McCall, Jiao, and Harris (2013) examined the structure of the MAP test for invariance, as a means of checking the construct validity of the MAP test. They noted several difficulties in the process. Specifically, they found the lack of availability of complete test bank data (98% of the test bank was missing for any given test) and the uniqueness of each individual test (complicating estimates of correlation) as problematic. Their strategy therefore was to investigate the data on the level of individual data clusters, such as Number Sense (on the Mathematics test). Their findings supported the validity findings previously published by the NWEA (2004).

The NWEA (2012) also published reports that investigated any predictive relationship between MAP and commonly used measures of college readiness, such as EXPLORE, PLAN and ACT. Findings from these studies indicated that students who sat the MAP exam and scored between the 70th and 80th percentile in reading and math, respectively, were achieving at the set benchmark for college readiness in these two areas. Scores between the 40th and 60th percentile on the MAP assessment were in line with the benchmark for college readiness in English language.

The MAP test shares several features with Smarter Balanced assessments in terms of its range of purpose, test design, and use of a single vertical scale to measure student growth.
The MAP test therefore reflects the most current trends in assessment and is designed to be taken several times per year, to provide up-to-date information on student achievement. Reports generated by MAP allow for interested teachers or administrators to see student achievement growth over time, and to consider the impact of the teacher on their growth. In this manner, investigations into what “value” a teacher adds to student growth can be made.

**Value-added models**

Value-added models (VAM) of student achievement arose as a next step to what are termed attainment models; namely, considering the effectiveness of a teacher based on a single-point-in-time measure of student achievement. In a study of the concept models used historically in attempts to determine teacher effects on student achievement, Rowan et al. (2002) concluded that the use of annual gains in VAMs was a more accurate and true method. In a review of the existing literature on VAMs, Koedel, Mihaly, and Rockoff (2015) wrote that policies should include information about value-added when it comes to making decisions on teaching personnel.

However, Kersting, Chen and Stigler (2013) cautioned that, while value added models tended to provide stable results for teachers who were considered to be in the top two-thirds in terms of performance, the results for teachers in the bottom one-third contained much more variance. As well, when testing several different VAM models, they noted that the use of different models could itself cause variation between 4% and 12% in terms of the classification of teacher effectiveness.

Additionally, Schochet and Chiang (2013) wrote that the measurement error for value added models can be as high as 26% for teachers who were considered to be average. This resulted in a likelihood that these teachers could be misclassified when decisions were taken based on student achievement results. Their findings for teachers whose students performed
higher or lower than the average had lower rates of measurement error, similar to Kersting, Chen and Stigler (2013).

Murphy (2012) noted several general difficulties with VAMs, including a lack of control for various effects (student, class, teacher, school, etc) that could impact the achievement of the student at that time. For example, if a student’s prior achievement was not accounted for (termed a student fixed-effect) nor if there was a variation in the demographics of students in a teaching group from one year to the next (termed a class-fixed effect). However, Kersting, Chen and Stigler (2013) tested different models, some of which controlled for class effects or student effects and others which controlled for prior achievement. They found that all the models were fairly consistent in the percentage produced that was associated with the teacher effect.

Murphy (2012) also questioned the comparisons resulting from value-added, noting that they may have unintended consequences because they are norm-referenced, such as causing teachers to not work collaboratively, for fear of jeopardizing their position in comparison to others. Braun (2005) wrote that no model can be comprehensive to draw causal inferences about a teacher’s effectiveness, and are therefore inherently inaccurate to some degree. Finally, Kersting, Chen and Stigler (2013) also noted that wide variations in class sizes can cause inaccuracy in the results of different VAM models. They suggested that minimum class sizes (around 15) only be considered.

There are several VAMs currently seen in the literature. As the nature of VAMs are to measure the growth in student achievement from one test to the next, one model that is used is simply termed the gains model. McCaffrey et al. (2003) noted that gains models typically use scores from spring-to-spring test dates, though they can also be used in fall-to-spring measurements as well. They define the model as follows:

\[ y_t - y_{t-1} = m_t + T_t + e_t \]
Where $t$ is the current year of testing, $y$ is the student achievement score, $m$ is the expected student average score (normed to the overall student population), $T$ is the teacher effect, and $e$ is a measure of residual error. In this model, other effects—whether they be at the school-level, student-level, or effects related to the prior year test score are not accounted for.

A strength of the gains model is that it focuses exclusively on student achievement (McCaffrey et al., 2003; Murphy, 2012; Rowan et al., 2002). However, Murphy (2012) and McCaffrey et al. (2003) noted the necessity of the student achievement variable to be measured on a single testing scale with a fixed determination of student growth from year to year. An additional strength is the more easily understood nature of growth, that can be presented to a wider audience in a non-technical manner.

A few limitations exist in the gains model as well, perhaps most significant is that differences amongst the students themselves are not accounted for (McCaffrey et al., 2003; Murphy, 2012). Additionally, as seen in Rowan et al. (2002), who tested several VAM models on their data set, Rogosa and Saner (1995) noted that a limitation occurs in this model when the amount of variance amongst student results is very small, which has the effect of under-representing teacher effect sizes on achievement. Finally, some studies (Buddin & Zamarro, 2009; Rivkin et al., 2005) have commented on the difficulty of measuring teacher and student effects; although it is somewhat mitigated through the use of successive data points, they found that these effects can impact the initial data point used.

A second VAM model is the Covariate Adjustment Model. This model differs from the gains model in that the current test score is measured based on its variation from the expected growth seen from the previous test score. This model is also termed a “residual effects” model (Murphy, 2012). Additionally, instead of treating the previous test score as fixed, it too is a function of the next prior test score. Hence, the model that measures current
achievement is still based on one variable (univariate), but then additional models are created based on the scores of all previous years (McCaffrey et al., 2003).

McCaffrey et al. (2003) wrote that a benefit of this model is the potential to treat the model not only in terms of projecting linear growth but that non-linear growth could be modeled as well. Thus, polynomial functions could also be developed. One limitation noted by McCaffrey et al. (2003) is that the covariate model actually looks at student achievement over time as in terms of the impact of several teachers, and not just the current teacher. They stress that this may be of value or interest, depending on the aim(s) of the research question(s). An additional limitation is the possible error resulting from lack of control of class selection, which may (likely) be non-random from year to year (Anderman, 2015).

In their analysis of the covariate model, Rowan et al. (2002) wrote that actual growth was not being measured; rather, the result was actually a measure of student achievement status (at a particular point in their educational career), while accounting for SES and level of previous achievement. However, Murphy (2012) noted that this result is a strength of this model, in that different characteristics of students are accounted for (in contrast to the gains model).

A third VAM is a multivariate longitudinal model. This model incorporates data from multiple years to simultaneously model student achievement for multiple years under study. As with the other models outlined previously, clarity on what exactly is being measured is necessary in the use of this model - in this case the aim is to investigate growth over time. Rogosa and Saner (1995) noted this shift in thinking to consider achievement as a “student growth curve” that may not be linear in nature. In the multivariate model, the teacher effect each year is considered to be linear and it remains from year to year; similar to the covariate model, the teacher effect is measured as a “deflection” from the average expected growth in the overall population under study (Braun, 2005; McCaffrey et al., 2003).
Although this model is generally seen to generate more reliable and complete measures of teacher effects on student growth (McCaffrey et al. 2003), it is not without limitations of its own. McCaffrey, Lockwood, Koretz, Louis, and Hamilton (2004) reviewed several studies that noted a change in studied content from year to year as having an impact on student achievement that was difficult to gauge. There also exists differing viewpoints on whether other variables should be controlled for, such as student demographics variables. In a description of one of the more commonly used multivariate models, EVAAS (Educational Value-Added Assessment System), Braun (2005) noted the debate as to whether the inclusion of controlled student variables would improve the accuracy of multivariate models. Sanders et al. (2009), whose work EVAAS is based upon, wrote that no correlation has been found between student demographics and achievement growth in several studies they conducted, hence the inclusion of the student effects is unnecessary.

Regardless of the choice of VAM, researchers have also investigated the value of the data gathered and what purpose it will serve. Kersting, Chen, and Stigler (2013) note that the choice of models must also depend on what conclusions researchers hope to draw, whether they be about individual teachers or groups of teachers. Kane and Steiger (2014) argued that the most accurate use and a more stable measure of VAM was to compare a teacher’s effectiveness in the current year to their career results and then consider their future expected performance. They note that this information is useful in helping school administrators draw comparisons between a current teacher’s performance and what they might expect from a replacement teacher.

Value-added models range in complexity and have been thoroughly researched. Nevertheless, several questions remain about their limitations and how to use them fairly when considering teacher impact on students’ achievement. Access to multiple years of data, the nature of the measure of student achievement (whether on a single vertical scale or not),
and whether controls can be put in place to account for various effects (school or student level) all impact the viability of the available VAMs. The study in this dissertation was therefore developed in light of these considerations.
Chapter 3

Methods

The nature of this study was quantitative as it aimed to examine the impact of teacher status on student achievement in mathematics in international schools by using numeric achievement data. For the purpose of this study, teacher status was measured by variables related to a teacher’s temporal experience in teaching, educational level, and country issuing their teaching certification. Student achievement in mathematics was collected from grades 3 to 8; achievement in mathematics was defined as the gain a student made between two predetermined testing sessions.

Research Questions

1. Does the overall number of years of teacher experience have an impact on student achievement in mathematics?

2. Does the number of years of teaching experience at the current school have an impact on student achievement in mathematics?

3. Does the location of the teacher’s certification have an impact on student achievement in mathematics?

4. Does the teacher’s level of education have an impact on student achievement in mathematics?

Setting & Participants

The sample data for this study was collected from international schools in Brazil who, as part of their methods of measuring student achievement, were measuring student the MAP test in Grades 3-8 to assess achievement in mathematics. International schools in Brazil that offer the MAP testing have somewhat similar student demographics; they tend to have high percentages of Brazilian nationals. Table 1 shows this percentage for the majority of schools that were contacted to participate in the study. These figures were publicly available on the
website of each school or in their public accreditation document. However, these percentages may mask inconsistencies in how nationality is determined; for example, it is not clear how multiple passport holders are reported across schools. Nevertheless, the majority national demographic in all international schools in Brazil is Brazilian, and by inference the percentage of students from other national backgrounds is also similar, making for reasonable comparisons across schools when using a common assessment.

Table 1
*Percentage of Brazilian students attending international schools.*

<table>
<thead>
<tr>
<th>School</th>
<th>Number of Brazilian students (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>School A</td>
<td>624 (51)</td>
</tr>
<tr>
<td>School B</td>
<td>493 (67)</td>
</tr>
<tr>
<td>School C</td>
<td>492 (50)</td>
</tr>
<tr>
<td>School D</td>
<td>413 (63)</td>
</tr>
<tr>
<td>School E</td>
<td>106 (31)</td>
</tr>
<tr>
<td>School F</td>
<td>271 (79)</td>
</tr>
<tr>
<td>School G</td>
<td>506 (68)</td>
</tr>
<tr>
<td>School H</td>
<td>279 (72)</td>
</tr>
<tr>
<td>School I</td>
<td>350 (74)</td>
</tr>
<tr>
<td>School J</td>
<td>724 (86)</td>
</tr>
<tr>
<td>School K</td>
<td>156 (52)</td>
</tr>
<tr>
<td>School L</td>
<td>342 (83)</td>
</tr>
</tbody>
</table>

The MAP test has seen increased use worldwide in international schools; in their 2013 spring report, the NWEA reported that over 100,000 students sat the MAP mathematics test outside of the United States. According to an international dashboard shared with its partner schools, the NWEA reported that this number increased to over 200,000 for the fall 2015/16 testing session. One of the reasons the MAP data was chosen for this study was because of how pervasive the use of this test was within the region. In the AASSA region, a total of
15,709 students took the mathematics achievement test for grades 3-8 in the fall 2015/16 session, an increase of 2.8% over the session the previous year. For the same time period and testing range, the number of students in Brazil who took the test was 3,554, an increase of 10%.

In summary, the selection of the MAP test for use in this study was appropriate based on its broad use and recent growth in South America and, more specifically, Brazil. It was estimated that there were 18-19 schools that were using the MAP testing during the 2015-16 school year (S. Blauer, personal communication – N.D). The researcher was able to find the contact details for 14 schools and approached all schools to participate in the study.

**Administration and data collection**

Schools were contacted at the end of the 2015-16 school year, approximately close to the time when they were completing their second (or last) round of MAP testing. The targeted teacher/class levels were from Grade 3 to Grade 8 of each school, provided that the MAP mathematics test was conducted at that year-level. A variety of levels were targeted to increase the sample size of student data and teacher participants.

For the schools that wished to participate, contact with a key school administrator was established to begin the data collection. Where possible, the researcher visited schools personally to ease the burden of data collection. There were three parts to collecting the necessary data for the study: determining the eligible teachers and their classes; collecting their class data; and collecting the teacher data.

**Determining eligible teachers and their classes**

To determine the eligible teachers and classes, the researcher and key administrator established how many classes were at each grade level, from 3rd grade to 8th grade. An organizing spreadsheet (Appendix A) was shared with the key administrator at each school that matched the teacher(s) at each grade level to a specified code. For example, if there were
two 6th grade classes in School A, they were denoted as A6.1, and A6.2. If a teacher taught at multiple grades or taught multiple classes of the same grade level, they were coded for each class group they taught in the specified range. Using the example above, the same teacher was coded for sections A6.1 & A6.2 if they taught both groups. After the organizing spreadsheet was completed, the key administrator kept access for use in coding the class data to the specific teacher.

**Collecting the class data**

Student achievement results by class were collected for the two sessions of MAP testing conducted that year, through the MAP report titled: *Achievement Status and Growth Summary Class Report (Fall 2015 - Spring 2016)*, hereafter known as the “Class Report.” The Class Report shows the achievement score for each student for the Fall 2015 and Spring 2016 testing session. The gain made between the testing sessions was calculated from these two data points.

For each Class Report, the key administrator of each school was asked to place the corresponding teacher code from the organizing spreadsheet in the top right-hand corner, and then remove the teacher name and student names from the Class Report to protect the anonymity of the teacher and students in the class. The resulting modified Class Report (Appendix B) was then returned to the researcher. This process was repeated for each class in the specified range (from Grades 3 – 8).

**Collecting the teacher data**

Teacher data was collected through the use of a Google survey (Appendix C). The researcher shared a copy of the survey with the key administrator of each school. The key administrator was asked to send out the survey to the teachers identified in Appendix A. The survey consisted of questions related to the teacher’s background, professional experience and qualifications. For the purposes of this study, the school year 2015-16 was counted
towards the total years of teacher duration and towards total years of teaching experience. This was because the school year had been mostly completed by the time the final round of MAP data was collected. To protect the anonymity of the teachers who responded to the survey, the researcher replaced the name of each teacher who responded to the survey by their corresponding code from Appendix 1. The researcher then assigned a number (from 1-38) for the presentation of the results, to ensure teachers would not be able to be identified by their professional characteristics.

**Analytical procedures**

Multiple regression was used to analyze the different relationships and potential strength of relationships between teacher status and student achievement. Osborne and Waters (2002) outlined four assumptions in multiple regression that must be examined for this statistical test to be used. They were: linearity, reliability of measurement, homoscedasticity and normality (p.1). Therefore, necessary steps were taken to address these assumptions.

As there is an assumption of linearity in multiple regression, residual plots for all continuous variables were checked for any non-linear features. To check for homoscedasticity, a studentized residual plot was checked; if heteroscedasticity was found, additional statistical tests were employed. A Durbin-Watson statistic was used to check for independence of observations. Residual plots were checked to see if residuals were normally distributed, and case-wise diagnostics were employed to identify any outliers beyond three standard deviations.

As there were multiple independent variables involving time (overall years of teaching experience & years of experience at current school), a phenomenon known as multicollinearity was of concern because the model might otherwise conflate the two different types of experience. Variance inflation factors for the continuous variables were
calculated to check for multicollinearity. If multicollinearity was found, one of the temporal variables was removed from the model.

A forced entry model was used for entering all variables into the regression equation. Hocking (2013) noted a preference for including all combinations of variables rather than taking a sequential (forward, backward, or stepwise) approach to developing a regression model. His caution in using this model was the level of computation required. SPSS and Stata statistical software were used to develop regression models and test regression assumptions. As there was no assumption of an order of importance amongst the independent variables, and current research has benefitted from powerful statistical software such as SPSS and Stata, a forced entry model was able to be used.

Variables

Several independent variables were used as measures of teacher status. As previously noted, ample research has been conducted to investigate if a relationship exists between a teacher’s overall number of years of teaching experience with student achievement (Buddin & Zamarro, 2009; Greenwald et al., 1996; Huang & Moon, 2008; Rivkin et al., 2005; Rowan et al., 2002) and a teacher’s level of education with student achievement (Brewer & Goldhaber, 1997; Huang & Moon, 2008; Rowan et al., 1997; Rowan et al., 2002). As these studies all examined teacher data from the United States, this study extended and built on this significant base of research and incorporated data from a context previously unexamined, by investigating teachers who were working in international schools.

The final independent variable was the location of where the teacher’s certification was issued. Unique to international schools like those found in Brazil is that they employ teachers from different national backgrounds. These teachers therefore earned their teaching certification according to the criteria determined by their respective national system, which can vary from country to country. The researcher was unable to find any studies that
investigated this phenomena. More generally, the current body of research on the topic of teacher certification is limited, confined to whether or not a teacher was certified in mathematics (Greenberg et al., 2004; Lubienski, Lubienski, & Crane, 2008) or whether a teacher was certified through a traditional or non-traditional manner (Shuls & Trivitt, 2015). Hence, this study provided an opportunity to investigate an aspect of teacher status that has been previously unexplored.

Results from the Measures of Academic Progress (MAP) mathematics test was the data source chosen for the dependent variable in this study. The dependent variable was represented by the gain made by a student in an academic calendar year on the MAP mathematics test. Gains scores were used as a measure of student achievement instead of actual achievement scores for several reasons. First, Rowan et al. (2002) noted the accuracy of a gains score model when investigating the relationship between teacher effects on student achievement. Additionally, Murphy (2012) and McCaffrey et al. (2003) wrote that a condition for using a gains score model is measurement on a single vertical scale, a condition which was met by the nature of the design of the MAP test. Finally, as previously noted, the collection of student achievement data in international schools is problematic. Collection of multi-year data posed an even greater challenge, thus eliminating the possible use of multi-year models.

As previously noted in Chapter 2, a limitation traditionally associated with a gains model is the lack of control for variation amongst students. Due to the regional nature of this study and the similarity in the composition of students in international schools in Brazil (both in terms of SES and nationality), this variation was controlled for, to a limited extent. However, the actual level of control for this variation was not quantified in the study.

The following independent and dependent variables were defined and operationalized for this study:
Independent variables

- Total years of teacher experience (hereafter known as experience) -- A discrete variable that included the current year being taught by the teacher.
- Total years of teaching experience at the current school (hereafter known as tenure) - A discrete variable that included the current year being taught by the teacher.
- Nature of teacher’s bachelor degree - A categorical variable that was operationalized in the following way:
  - In mathematics
  - In different discipline
  - Does not have
  - Similar to a bachelor’s degree.
- Nature of teacher’s master’s degree - A categorical variable that was operationalized in the following way:
  - In mathematics
  - In different discipline
  - Does not have
  - Similar to a master’s degree.
- Geographical location of the teacher’s teaching certification - A categorical variable that was operationalized in the following way:
  - Earned certification in Brazil
  - Earned certification in United States
  - Earned certification in another country

Dependent variable
• Mathematics achievement - A discrete variable that was operationalized as the gain made by a student from their the Fall 2015 MAP mathematics test score to their Spring 2016 MAP mathematics test score.

Research Ethics

The anonymity of student and teacher data was an integral part of this study, as the intent of the study was not to gather evaluative data about specific teachers or measure specific growth of individual students. Rather, it was to investigate relationships between groups of teachers who share similar demographic characteristics and their overall student results. Hence, it was imperative that the design of the study be conducted in a manner that protected individual anonymity. Teachers provided consent to the use of their data (that was kept anonymous) through their voluntary participation in the survey—they were not required by their school or the researcher to participate. As personal information on students was not collected, it was unnecessary to have the consent of individual students. The methodological design of the study was submitted to and approved by the George Fox University Institutional Review Board (IRB).

At the time of the study, the researcher was a graduate student at George Fox University, and had a strong connection both to Brazil and the subject of mathematics for the past 15 years. As such, it was important to be particularly mindful of any bias or expectation with the results found in the analysis of data. The researcher therefore endeavoured to be cautious throughout the process of data collection and subsequent analysis of any personal expectation (from the researcher or the participating administrator), both conscious or otherwise. For example, the student achievement data was gathered after the testing sessions had been completed, so that there would be no possibility for the data to be influenced as a result of a school’s participation in the study.
Several measures were taken to protect the anonymity of the data while it was in the possession of the researcher. After the participating administrator provided the names of the possible teacher participants, their access to the shared file (Appendix A) was removed. As well, the participating administrator did not have access to the survey results provided by the teachers. Lastly, when creating the summary data to produce descriptive statistics and the multiple regression analysis, the researcher organized the teacher data in a manner that any reader of this study would not be able to identify a specific teacher.
Chapter 4

Results

The results of the collection of student and teachers were then used to proceed to the analysis stage of the study. Descriptive statistics were calculated to provide an overall picture of the participants. Assumptions of multiple regression were tested to determine whether or not a multiple regression model could be developed. Finally, a series of analyses were performed to investigate possible findings from the data.

Fourteen international schools in Brazil that utilized MAP to measure mathematics achievement were contacted to solicit their participation in the study. Eight schools initially responded with interest. Over the period of time that data was collected, one school withdrew due to a decision to no longer use the MAP test. Therefore, seven schools were able to successfully participate in the study by administering the teacher survey and sharing the teacher and student data with the researcher.

Descriptive statistics of participants

Of the seven schools, 59 teachers were identified for possible participation, as they were teachers of one (or more) mathematics classes from grades 3-8. In communication with an administrator from each school, the researcher checked that the identified teachers were suitable to be included in the study. There was one criterion for being included in the study, that the same teacher teach the class for the entire school year, covering the first MAP testing session to the second. As a result of this check, the number of potential teacher participants of certain schools decreased, due to different factors.

Several teachers from one particular school were removed because they had not been the teacher of their grade level for the entire school year. From a different school, a team teaching approach was used at each grade level. The researcher elected to include responses from this school under two conditions. The first condition was that both teachers of a
particular grade level had to have provided answers to the questionnaire. This condition nullified one potential participant in the study. The second condition was that the data collected from the pair of team teacher responses was included only for each question(s) in which they gave an identical answer(s). The team teaching pair were then included in the study as single teacher data points for each grade level.

Finally, several teachers elected not to respond to the survey, as it was voluntary. In total, the researcher considered there to be 38 unique teacher survey responses that were suitable for use out of the possible 59 teachers identified. There were 1289 students in the classes of these 38 teachers, who took the MAP test during both the Fall 2015 and Spring 2016 testing sessions. Each student earned a measure of growth (either positive or negative) between the testing sessions. Thus, there were 1289 individual data points that constituted the set of data for the dependent variable in this study. Table 2 shows the summary of teacher responses to the survey questions, detailing the professional profile of each teacher. The profile of teachers 36, 37 and 38 represent the individual data points for the team – teaching partners mentioned above; the fields denoted by an asterisk were not included in this study.

Table 2
Teacher demographics

<table>
<thead>
<tr>
<th>Teacher code</th>
<th>Degrees</th>
<th>Years of experience</th>
<th>Tenure</th>
<th>Location of certification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bachelor (Other), Master (Other)</td>
<td>8</td>
<td>2</td>
<td>U.S.</td>
</tr>
<tr>
<td>2</td>
<td>Bachelor (Other, Master (Other)</td>
<td>16</td>
<td>11</td>
<td>Brazil</td>
</tr>
<tr>
<td>3</td>
<td>Bachelor (Other), Master (Other)</td>
<td>32</td>
<td>22</td>
<td>U.S.</td>
</tr>
<tr>
<td>4</td>
<td>Bachelor (Other), Master (Other)</td>
<td>10</td>
<td>3</td>
<td>U.S.</td>
</tr>
<tr>
<td>5</td>
<td>Bachelor (Math), Master (Other)</td>
<td>4</td>
<td>2</td>
<td>Brazil</td>
</tr>
<tr>
<td>6</td>
<td>Bachelor (Other), Master (Other)</td>
<td>8</td>
<td>6</td>
<td>Brazil</td>
</tr>
<tr>
<td>7</td>
<td>Bachelor (Other), No Master</td>
<td>6</td>
<td>1</td>
<td>U.S.</td>
</tr>
<tr>
<td>8</td>
<td>Bachelor (Other), Master (Other)</td>
<td>12</td>
<td>2</td>
<td>U.S.</td>
</tr>
<tr>
<td>9</td>
<td>Bachelor (Other), Master (Other)</td>
<td>4</td>
<td>4</td>
<td>U.S.</td>
</tr>
<tr>
<td></td>
<td>Degree Combination</td>
<td>Count 1</td>
<td>Count 2</td>
<td>Country</td>
</tr>
<tr>
<td>----</td>
<td>------------------------------------------</td>
<td>---------</td>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>10</td>
<td>Bachelor (Other), Master (Other)</td>
<td>8</td>
<td>6</td>
<td>U.S.</td>
</tr>
<tr>
<td>11</td>
<td>Bachelor (Math), No Master</td>
<td>21</td>
<td>1</td>
<td>Brazil</td>
</tr>
<tr>
<td>12</td>
<td>Bachelor (Other), Master (Other)</td>
<td>12</td>
<td>3</td>
<td>U.S.</td>
</tr>
<tr>
<td>13</td>
<td>Bachelor (Other, Similar to Master)</td>
<td>10</td>
<td>5</td>
<td>Brazil</td>
</tr>
<tr>
<td>14</td>
<td>Bachelor (Other), Master (Other)</td>
<td>12</td>
<td>2</td>
<td>Other</td>
</tr>
<tr>
<td>15</td>
<td>Bachelor (Other), Master (Other)</td>
<td>30</td>
<td>22</td>
<td>Other</td>
</tr>
<tr>
<td>16</td>
<td>Bachelor (Other), Master (Other)</td>
<td>9</td>
<td>7</td>
<td>U.S.</td>
</tr>
<tr>
<td>17</td>
<td>Bachelor (Other), No Master</td>
<td>3</td>
<td>1</td>
<td>U.S.</td>
</tr>
<tr>
<td>18</td>
<td>Bachelor (Other), No Master</td>
<td>47</td>
<td>47</td>
<td>Brazil</td>
</tr>
<tr>
<td>19</td>
<td>Bachelor (Other), Master (Other)</td>
<td>3</td>
<td>2</td>
<td>Other</td>
</tr>
<tr>
<td>20</td>
<td>Bachelor (Other), Master (Other)</td>
<td>24</td>
<td>6</td>
<td>Brazil</td>
</tr>
<tr>
<td>21</td>
<td>Bachelor (Other), Master (Other)</td>
<td>12</td>
<td>7</td>
<td>U.S.</td>
</tr>
<tr>
<td>22</td>
<td>Bachelor (Other), Master (Other)</td>
<td>37</td>
<td>4</td>
<td>U.S.</td>
</tr>
<tr>
<td>23</td>
<td>Bachelor (Math), No Master</td>
<td>20</td>
<td>12</td>
<td>U.S.</td>
</tr>
<tr>
<td>24</td>
<td>Bachelor (Other), No Master</td>
<td>17</td>
<td>1</td>
<td>U.S.</td>
</tr>
<tr>
<td>25</td>
<td>Bachelor (Math), Master (Other)</td>
<td>16</td>
<td>16</td>
<td>Brazil</td>
</tr>
<tr>
<td>26</td>
<td>Bachelor (Other), Master (Other)</td>
<td>19</td>
<td>19</td>
<td>Brazil</td>
</tr>
<tr>
<td>27</td>
<td>Bachelor (Other), No Master</td>
<td>1</td>
<td>1</td>
<td>Brazil</td>
</tr>
<tr>
<td>28</td>
<td>Bachelor (Math), Master (Math)</td>
<td>15</td>
<td>12</td>
<td>Brazil</td>
</tr>
<tr>
<td>29</td>
<td>Bachelor (Other), Master (Other)</td>
<td>27</td>
<td>27</td>
<td>Brazil</td>
</tr>
<tr>
<td>30</td>
<td>Bachelor (Other), Master (Other)</td>
<td>15</td>
<td>1</td>
<td>U.S.</td>
</tr>
<tr>
<td>31</td>
<td>Bachelor (Other), Master (Other)</td>
<td>12</td>
<td>1</td>
<td>U.S.</td>
</tr>
<tr>
<td>32</td>
<td>Bachelor (Other), Master (Math)</td>
<td>7</td>
<td>2</td>
<td>U.S.</td>
</tr>
<tr>
<td>33</td>
<td>Bachelor (Math), No Master</td>
<td>18</td>
<td>2</td>
<td>U.S.</td>
</tr>
<tr>
<td>34</td>
<td>Bachelor (Other), Master (Math)</td>
<td>18</td>
<td>1</td>
<td>U.S.</td>
</tr>
<tr>
<td>35</td>
<td>Bachelor (Other), Master (Other)</td>
<td>20</td>
<td>8</td>
<td>U.S.</td>
</tr>
<tr>
<td>36</td>
<td>Bachelor (Other), No Master</td>
<td>**</td>
<td>**</td>
<td>Brazil</td>
</tr>
<tr>
<td>37</td>
<td></td>
<td>**</td>
<td>20</td>
<td>Brazil</td>
</tr>
<tr>
<td>38</td>
<td>Bachelor (Math)</td>
<td>**</td>
<td>**</td>
<td>Brazil</td>
</tr>
</tbody>
</table>
Table 3 shows the survey responses provided by the sampled teachers. The results indicated that participating teachers were very experienced, though many were relatively new to their current school. This finding appeared to be consistent with the elevated rates of teacher turnover in international schools. Due to the relatively low number of teachers in the sample, there were low numbers of teachers who possessed bachelor’s and/or master’s degrees in mathematics. This may have been impacted by two different factors. First, as half of the targeted grades were in the elementary school, having a specialized degree in mathematics was likely to be uncommon. Second, there were fewer numbers of middle school teachers who responded to the survey, as these teachers often taught multiple sections in different grade levels, whereas at the elementary level there was one teacher per section. So elementary teachers were overrepresented in comparison with middle school teachers. As many of the schools that participated in the study offered an American curriculum and were accredited by an accrediting body from the United States, it was to be expected that the majority of teacher participants received their certification either in Brazil or in the United States.

Table 3
Teacher survey responses, including corresponding average student achievement growth

<table>
<thead>
<tr>
<th>Response</th>
<th>Frequency</th>
<th>N*</th>
<th>Average Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A.1 Total years of teaching experience</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 or less</td>
<td>5</td>
<td>115</td>
<td>10.5</td>
</tr>
<tr>
<td>More than 5</td>
<td>31</td>
<td>1020</td>
<td>8.28</td>
</tr>
<tr>
<td><strong>A.2 Total years teaching at current school (tenure)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 4</td>
<td>17</td>
<td>608</td>
<td>8.43</td>
</tr>
<tr>
<td>4 - 10</td>
<td>9</td>
<td>251</td>
<td>8.86</td>
</tr>
<tr>
<td>11 or above</td>
<td>9</td>
<td>251</td>
<td>8.83</td>
</tr>
<tr>
<td><strong>A.3 Nature of bachelor's degree</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathematics</td>
<td>7</td>
<td>82</td>
<td>4.5</td>
</tr>
<tr>
<td>Not Mathematics</td>
<td>30</td>
<td>928</td>
<td>9.13</td>
</tr>
<tr>
<td><strong>A.4 Nature of master's degree</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathematics</td>
<td>3</td>
<td>304</td>
<td>7.2</td>
</tr>
</tbody>
</table>
Figures 1-3 (Appendix D) provide visual displays of the three categorical variables (bachelor’s degree, master’s degree & country of certification) and their relationship with student achievement growth. As can be observed in Table 3, there were answer categories that were rarely chosen by the teacher participants, or not at all. For instance, all teachers who responded in the survey possessed a bachelor’s degree, thus negating the possible answers of “does not have” or “similar degree to a bachelors,” which can be observed in Figure 1. As well, only three respondents possessed master’s degrees in mathematics, and only three respondents reported that they earned their teaching certification from a country outside of Brazil or the United States.

Despite the imbalances observed in the reported answers to categorical questions in Table 3, Figures 1-3 show a generally similar trend in the range of student achievement growth across the teacher survey responses. However, this interpretation should be treated with caution as in some cases the box plot represents student achievement growth results that are connected to few (less than 5) teachers.

Digging further into the data, Table 4 shows a breakdown by grade level of the teacher participants in the study. Several teachers at the middle school level taught multiple sections or grade levels of mathematics, hence the number of unique participating teachers was low in these grades while the number of student data points was still close to that of other grade levels. Table 4 therefore over reports (by eight) the actual number of teacher
participants in the study. For example, in one school the same teacher taught all students (in multiple sections) from grades 6, 7 and 8. A pattern of higher levels of growth in the elementary years when compared to the middle school years is immediately evident in Table 4.

Table 4  
*Average student achievement growth by grade level*

<table>
<thead>
<tr>
<th>Grade</th>
<th># of teachers</th>
<th># of students</th>
<th>Average Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 3</td>
<td>8</td>
<td>201</td>
<td>11.6</td>
</tr>
<tr>
<td>Grade 4</td>
<td>11</td>
<td>241</td>
<td>10.3</td>
</tr>
<tr>
<td>Grade 5</td>
<td>8</td>
<td>174</td>
<td>9.60</td>
</tr>
<tr>
<td>Grade 6</td>
<td>7</td>
<td>243</td>
<td>5.89</td>
</tr>
<tr>
<td>Grade 7</td>
<td>6</td>
<td>214</td>
<td>7.63</td>
</tr>
<tr>
<td>Grade 8</td>
<td>5</td>
<td>188</td>
<td>4.68</td>
</tr>
</tbody>
</table>

**Testing assumptions of multiple regression**

Before the regression model could be developed, the assumptions for multiple regression were verified for the data. Outliers were investigated first to see if all of the data could be considered. Case-wise diagnostics were employed to test for outliers: 1.3% of the data was found to be outside of three standard deviations. As this percentage was so low, the researcher elected to include all data in the analysis.

Linearity of the continuous variables (experience & tenure) with the dependent variable (student achievement growth) was the first assumption to be tested. Figures 4 and 5 (Appendix E) show the partial plots of these individual continuous independent variables. As can be seen from the figures, no curvilinear pattern could be observed for either independent variable. Hence, the assumption of linearity was considered to have been met, a precondition for continuing with the multiple regression analysis.
To test for homoscedascity, a studentized residual plot was examined to determine whether or not the data was evenly distributed. The result is Figure 6 (Appendix E). As no funnel shape can be observed from the figure, the assumption of homoscedasticity was assumed to have been met as well.

To test for the assumption of normality of the continuous independent variables, a P-P plot was run. The result is Figure 7 (Appendix E). As can be observed from the plot, the residuals were very close to the best-fit line. This showed that the error between the predicted values and the actual values fell within a normal range, which is needed for multiple regression analysis.

Finally, a Durbin-Watson statistic was calculated to check for independence of observations. The calculation yielded a value of 1.97, indicating that the observations were indeed independent. Hence, all the conditions necessary for multiple regression were considered to have been met. However, since there were multiple temporal independent variables, the possible effect of multicollinearity needed to be investigated.

To check for multicollinearity, Table 5 shows the Variance Inflation Factors (VIF) for the independent (continuous) variables. As the values of 1/VIF were less than 10, these variables were not considered to be collinear. This meant that the numerical variables (experience and tenure) did not significantly impact each other in the study, and both variables could continue to be used.

Table 5

<table>
<thead>
<tr>
<th>Variable</th>
<th>VIF</th>
<th>1/VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>experience</td>
<td>2.95</td>
<td>0.339</td>
</tr>
<tr>
<td>tenure</td>
<td>2.29</td>
<td>0.437</td>
</tr>
</tbody>
</table>
**Multiple regression analysis**

With the assumptions of multiple regression having been met and the absence of multicollinearity, the researcher could continue to build the regression model. Table 6 shows the results of the first regression analysis that was performed. A forced-entry approach was used to incorporate all variables, using dummy variables for those that were categorical. In this first analysis, several variables were found to have a statistically significant relationship with student achievement growth, though the overall model was able to explain only 5% of the variance in student achievement growth ($R^2 = 0.054$).

For the continuous variables, experience was not shown to be statistically significant, while tenure was. As the standard deviations for growth ($SD = 7.08$) and tenure ($SD = 7.68$) were quite similar, the impact of tenure was a low negative value. Hence, for each additional year of tenure, the amount of student achievement growth decreased by one-eighth of a point, on average.

For the categorical variables, teachers whose bachelor’s degrees and master’s degrees were not in the field of mathematics had a statistically significant, strong positive impact (one third of a standard deviation) on student growth. Lastly, there was a strong negative impact (almost half of a standard deviation) on growth for teachers whose certification was from the United States or another country outside of Brazil. However, this initial was quite limited as it did not incorporate the clear trend of decreased growth in higher grade levels that can be observed in Table 4.
Table 6  
*Regressed growth against all variables*

| Variable | Coef. | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|----------|-------|-----------|-------|------|----------------------|
| experience | 0.009 | 0.041     | 0.220 | 0.824 | -0.072 - 0.091 |
| tenure   | -0.120 | 0.052     | -2.32 | 0.020 | -0.222 - -0.019 |
| Bachelor other | 2.67  | 0.653     | 4.09  | 0.000 | 1.39 - 3.95 |
| Master other | 2.64  | 0.584     | 4.51  | 0.000 | 1.49 - 3.78 |
| none     | 0.360  | 0.687     | 0.520 | 0.600 | -0.989 - 1.71 |
| similar  | -0.931 | 1.87      | -0.500 | 0.619 | -4.60 - 2.74 |
| Certification U.S. | -3.33 | 0.687     | -4.85 | 0.000 | -4.68 - -1.98 |
| Other    | -3.30  | 1.21      | -2.72 | 0.007 | -5.68 - -0.92 |
| _cons   | 8.08   | 0.822     | 9.84  | 0.000 | 6.47 - 9.69 |

Note: adjusted $R^2 = 0.0536$, $N = 1110$

As Table 4 showed a clear negative linear trend between grade and student achievement, Table 7 showed the same initial regression analysis as Table 6, however this time incorporating differences in grade. This model was now able to explain 11% of the variation in student achievement growth ($R^2 = 0.1152$). In this analysis, the variable for experience now had a low positive, but statistically significant, relationship with growth. Tenure and bachelor’s degree other than math continued to have an impact that was statistically significant. The impact of tenure was slightly more negative than in the previous model, whereas the impact of another bachelor’s degree was less positive than before. This indicated that the majority of teachers holding bachelor’s degrees in mathematics taught at the upper grade levels of the study. The impact of a master’s degree was also less positive in comparison with the first model, and was no longer significant at the 95% confidence level.

The negative impact of a teacher’s certification being from the United States was halved in this second analysis. This meant that the number of teachers with certification from the United States was disproportionally represented in the higher grade levels. The impact of other country’s certification continued to be largely negative, and still statistically significant.
Table 7

Regressed growth against all variables, including grade

| Variable          | Coef. | Std. Err. | t    | P>|t|  | 95% Conf. Interval |
|-------------------|-------|-----------|------|------|-------------------|
| experience        | 0.091 | 0.041     | 2.21 | 0.027 | 0.010 - 0.172     |
| tenure            | -0.193| 0.051     | -3.80| 0.000 | -0.292 - -0.093   |
| Bachelor other    | 1.41  | 0.647     | 2.17 | 0.030 | 0.137 - 2.68      |
| Master other      | 0.900 | 0.598     | 1.50 | 0.133 | -0.274 - 2.07     |
| Master none       | -1.68 | 0.704     | -2.39| 0.017 | -3.07 - -0.303    |
| Master similar    | -2.80 | 1.82      | -1.54| 0.124 | -6.37 - 0.770     |
| Certification U.S.| -1.44 | 0.698     | -2.07| 0.039 | -2.81 - -0.076    |
| Certification Other | -3.07 | 1.17 | -2.62 | 0.009 | -5.37 - -0.768    |
| grade             | -1.53 | 0.173     | -8.82| 0.000 | -1.87 - -1.19     |
| _cons             | 16.9  | 1.28      | 13.2 | 0.000 | 14.4 - 19.4       |

Note: adjusted $R^2 = 0.115$, $N = 1110$

The final regression analysis performed incorporated changes to the categorical variables involving master’s degree and country of certification. Because the survey question involving the master’s degree yielded low numbers of responses for those who held a degree in mathematics or a similar degree, the researcher re-grouped the responses to compare those who had a master’s degree (or similar) and those who did not. Similarly, responses to the questions of country of certification were re-grouped to those who were certified in Brazil and those who were certified internationally. Table 8 showed the results of this regression analysis, while continuing to factor in the impact of the grade level taught.

In this analysis, the impact of experience and tenure remained the same, both yielding a significant relationship while having a very small impact. The impact of having a master’s degree (regardless of the nature of the degree) was positive and statistically significant. The impact of having international certification was still negative, though this value was not significant at the 95% threshold.
Table 8
*Regressed against all variables, including combined variables “has Master” and “International Certification”*

| growth          | Coef.  | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|-----------------|--------|-----------|-------|------|---------------------|
| experience      | 0.085  | 0.041     | 2.07  | 0.039| 0.004 - 0.165       |
| tenure          | -0.172 | 0.049     | -3.48 | 0.001| -0.268 - -0.075     |
| Bachelor        | 1.38   | 0.643     | 2.14  | 0.033| 0.115 - 2.64        |
| other           |        |           |       |      |                     |
| Master          | 2.01   | 0.611     | 3.30  | 0.001| 0.815 - 3.21        |
| has             |        |           |       |      |                     |
| Certification   | -1.22  | 0.654     | -1.86 | 0.064| -2.50 - 0.069       |
| International   |        |           |       |      |                     |
| grade           | -1.55  | 0.160     | -9.70 | 0.000| -1.87 - -1.24       |
| _cons           | 15.2   | 1.05      | 14.4  | 0.000| 13.1 - 17.3         |

Note: adjusted $R^2 = 0.111$, $N = 1110$
Chapter 5

Discussion & Conclusions

The intent of this research study was two-fold: to extend prior research on the relationship between teacher professional characteristics and student achievement by applying it in a new context (in the setting of international schools), and to add to the overall base of research that has been undertaken in international schools. Both objectives were accomplished to a certain extent, though challenges presented themselves along the way. The mindset of the researcher when conducting the study was that the two objectives were inextricably linked. Difficulties that are inherent in meeting the latter objective both influenced and created a limitation for the potential success of achieving the former.

As anticipated, collecting data from multiple international schools proved to be very challenging. Common assessment measures and a culture of sharing or pooling data are not practices that are currently in place amongst schools of this nature, though the prevalence of use of MAP data in the region did assist. Without an existing data set, the researcher’s rate in successfully soliciting the participation of schools and teachers was impaired, resulting in a data set that was smaller than what could have been gathered. Successfully gathering data from schools and teachers who were willing to participate in the first instance was also a challenge; the process of data collection occurred over a much longer period of time, quite understandably due to the low priority given by the participating schools relative to their normal functions. Complications or challenges of these types likely explain why the current body of research on international schools is so small.

The complications in data collection experienced by the researcher resulted in a low number of responses obtained for certain questions, which limited the extent to which an analysis could be conducted. For example, in the final regression analysis, exploring the significance of a teacher having a master’s degree in mathematics was not possible due to the
small number of teacher participants who held this degree. This was similarly true for teacher participants whose teacher certification was obtained in a country other than Brazil or the United States. Nevertheless, a regression analysis of the data collected could be performed.

**Research questions – results & discussion**

The following research questions were posed in this study:

1. Does the overall number of years of teacher experience have an impact on student achievement in mathematics?

2. Does the number of years of teaching experience at the current school have an impact on student achievement in mathematics?

3. Does the location of the teacher’s certification have an impact on student achievement in mathematics?

4. Does the teacher’s level of education have an impact on student achievement in mathematics?

To answer the first research question, the final regression analysis yielded a very small, positive relationship between teacher experience and student achievement growth. While this result was statistically significant, the size of its impact was relatively insignificant, translating to roughly one-twelfth of a point in additional student achievement growth for each year of experience that a teacher possesses. This result generally matches with the findings of previous research (Hanushek, 1997; Rivkin et al., 2005; Rowan et al., 2002) on teacher experience.

The results of the second and third research questions were of particular interest as they delved into some of the unique characteristics of international schools. They are workplaces for teachers of different nationalities and educational backgrounds, and they may have elevated rates of teacher turnover. A logical inference may be made that teachers who have training from other countries outside of Brazil may have lower levels of tenure, due to
their status as an expatriate teacher with an option to return to their home country. This was indeed the case in this study; the average tenure of teachers with training from the United States was 4.55 years. The average tenure for Brazilian teachers was 9.72 years (not including one outlier data point, a teacher with 47 years of tenure at the school).

When considering the issue of tenure, it can be hypothesized that having teachers who stay at an international school for an extended period of time would be beneficial. In the case of locally-contracted teachers, they may experience a sense of connection and community with the students and the school. For the expatriate teacher, if they can navigate the adjustment that occurs when living in a new culture and stay for an extended period of time, they may have a higher level of appreciation for and connection with the new culture.

The impact of tenure turned out to be negative, though small. One might assert that international schools will benefit from teachers who stay longer; these findings do not lend support to this assertion. The negative impact largely offsets the mildly positive impact of teacher experience, rendering the results of both variables to be fairly inconclusive. Perhaps a wider-ranging study drawing on a large sample of teachers might yield more insights into the interesting question of tenure in international schools.

The regression analysis that had the most helpful information when considering the third research question is found in Table 7. To begin, however, the re-grouping of this question into teachers who were trained in Brazil versus teachers who earned their training internationally did not provide any additional insights, and the strength of the overall model was below the 95% confidence interval. Therefore, with such low numbers of teachers with certification from other countries, the most interesting comparison to investigate was between teachers whose training occurred in Brazil or in the United States.

It was evident from Tables 6 and 7 that teachers who were certified in Brazil had higher levels of student achievement growth than their counterparts from the United States;
however, this disparity lessened when accounting for grade level. Looking at the raw data for this question, a disproportionate number (12 out of 15) of Brazilian teachers who responded to the survey were teachers in the elementary year levels (grades 3–5), whereas the figures for teachers with the training from the United States were evenly split, with half the teachers (10) being in these early year levels.

Hence, a comparison can be drawn across elementary year levels as there was relatively equal representation of Brazilian-trained teachers and United States-trained teachers. However, a similar comparison could not be drawn at the level of middle school (grades 6–8). Only three Brazilian teachers in the sample taught at these grades in comparison to 10 teachers with training from the United States. This lack of data in particular year levels from teachers of different nationalities limits drawing conclusions; future studies will benefit from accessing a larger data set.

Nevertheless, that Brazilian teachers had marginally higher student achievement results may be a surprising result and worthy of additional research. The high level of growth in the overall number of international schools worldwide would seem to indicate a demand for teachers who come from countries that maintain a perception of having strong teacher training programs, such as those in the United States. Gillies (2001) notes the high level of demand for international schools and the generally high level of quality of expatriate teachers. In the experience of the researcher, it is taken for granted in international schools in Brazil that the teacher training one would experience in the United States would be superior. It is interesting to note that the results of this study do not support that perception. It is also interesting to see that the results connected to research questions 2 and 3 are at odds with each other—tenure having a low negative impact on student achievement growth while earning teacher certification in Brazil having a positive impact.
The fourth research question, related to the impact of level of education on student achievement growth, yielded interesting results when examining the data. Teachers possessing a bachelor’s degree in a subject other than mathematics had higher levels of achievement growth, similar to the findings of Rowan et al. (2002). Comparisons of type of master’s degree (mathematics versus other) could not be made from this data set, as the number of teachers with master’s degrees in mathematics was extremely low. Teachers who had a master’s degree (regardless of specialization) had slightly higher levels of achievement growth than teachers who did not; this finding is not consistent with many studies (Buddin & Zamarro, 2009; Greenwald et al., 1996; Hanushek, 1997; Huang & Moon, 2008; Rivkin et al., 2005).

**Limitations of the study**

The findings above should be treated with caution for several reasons. As previously mentioned, the small sample size in the study meant that several different aspects could not be fully explored as intended. As well, since the regression equation only accounts for 11% of the change in student achievement growth, there is a limit to the strength of conclusions that can be drawn. Some limitations not previously identified in the design of the study also appeared as the teacher data was collected.

One limitation was related to question design and interpretation of the survey; as the teacher’s self-reported their responses, there was the possibility of participant error in answer selection (the option to edit their response was not given) or question vagueness (for example, if a teacher obtained certification in more than one country). Another limitation was the language used in the administration of the survey. Although it was assumed that all the participating teachers could understand and respond to a survey given in English, for many of the participants English was not their first language. Hence, it was possible that some participants misunderstood or misinterpreted a question prompt due to language.
An additional limitation of the study was the lack of control for teacher or student factors. This has been identified in the research as a limitation of the use of a gains model (McCaffrey et al., 2003; Murphy, 2012). There may have been unique aspects in international school settings that could not be accounted for in this study, such as cultural differences when teachers and students come from different backgrounds. The effects of teachers from different nationalities is an interesting topic that is worthy of future exploration, but any impact of these differences was not accounted for in the study.

**Implications for future research**

Overcoming the barrier of data collection in international schools will continue to pose a significant challenge for quantitative researchers. A way forward may be for associations of international schools like AASSA to develop protocols for yearly data collection and sharing amongst its member schools. For associations in which specific assessment tools are already commonly used by a large number of schools, this step may not be too onerous. Organizations such as the NWEA could play a role in facilitating this exchange, while vigilantly maintaining the necessary level of anonymity in the data. Additionally, accreditation agencies like CIS may consider its role in the collection of survey data and how it can be shared to the benefit of the international school community for future research.

Extending current research further on mathematics teachers and the relationship between their professional characteristics and student achievement growth will also be problematic for the reasons previously mentioned. In addition, current assessments at the high school level in international schools do not currently utilize single vertical scales to measure growth. Instead, the CIE IGCSE exams and the IB Diploma exams measure achievement at a “single point in time.” Hence, value-added growth models cannot be considered. Were these exam boards to administer surveys to teachers in addition to the student achievement data
they receive from the exams, analysis could be conducted along the lines of the research studies conducted by Klecker (2002) and Greenberg et al. (2004).

Further exploration of the impact locally contracted staff and expatriate staff have on student achievement is also worthy of pursuit. Odland & Ruzicka (2009) note in their study the wide disparity in salary and benefits between these two groups of teachers. The findings of this study would seem to indicate that, based on student achievement results, this disparity may not be justified.

The documented growth (Brummitt, 2009) and popularity of international schools (Gillies, 2001) is evidence of the important role these schools play in forming young scholars for the future. A focus on teacher quality and its impact on student achievement is therefore a priority for schools who wish to understand and measure their own success. Finding ways to collect and use quantitative data in this endeavor is necessary, and consideration must be given as to how to overcome the current obstacles standing in the way.
References


NWEA. (2015). *Your introduction to MAP from NWEA*. Retrieved from: 


Rowan, B., Correnti, R., & Miller, R. (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


The Bill & Melinda Gates Foundation. (2013). *Ensuring fair and reliable measures of effective teaching.* Retrieved from:


Wang, S., McCall, M., Jiao, H., & Harris, G. (2013). Construct validity and measurement invariance of computerized adaptive testing: Application to measures of academic progress (MAP) using confirmatory factor analysis. *Journal of Educational and Developmental Psychology, 3*(1), 88-100. doi:


Appendices
Appendix A: Sample data collection spreadsheet

<table>
<thead>
<tr>
<th>Teacher Code</th>
<th>Teacher name</th>
<th>Teacher Code</th>
<th>Teacher name</th>
<th>Teacher Code</th>
<th>Teacher name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curt.3.1</td>
<td></td>
<td>Curt.3.2</td>
<td></td>
<td>Curt.3.3</td>
<td></td>
</tr>
<tr>
<td>Curt.4.1</td>
<td></td>
<td>Curt.4.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curt.5.1</td>
<td></td>
<td>Curt.5.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curt.6.1</td>
<td></td>
<td>Curt.6.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curt.7.1</td>
<td></td>
<td>Curt.7.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curt.8.1</td>
<td></td>
<td>Curt.8.2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B: Sample ASG class report

<table>
<thead>
<tr>
<th>Mathematics</th>
<th>ASG Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achievement Status</td>
<td>Growth</td>
</tr>
<tr>
<td></td>
<td>Fall 2016</td>
</tr>
<tr>
<td>PAIS Grade</td>
<td>PAIS Date</td>
</tr>
<tr>
<td>4</td>
<td>6/1/16</td>
</tr>
<tr>
<td>4</td>
<td>6/1/16</td>
</tr>
<tr>
<td>4</td>
<td>6/1/16</td>
</tr>
<tr>
<td>4</td>
<td>6/1/16</td>
</tr>
</tbody>
</table>

Summary for Mathematics

<table>
<thead>
<tr>
<th>Percentage of Students who Met or Exceeded the Projected RT</th>
<th>Percent of Projected Growth Met</th>
<th>Count of Students who Met or Exceeded the Projected RT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Explanatory Notes

Generated: 1/30/16, 1:07:32 PM
Appendix C: General teacher survey

Survey - Teacher Profiles

Dear teacher,

Thank you in advance for filling out this brief survey. As part of my dissertation work, I am investigating possible relationships between student achievement and international school mathematics teachers' professional experience profiles.

My research is structured so that information on teachers and students are anonymous. This survey is designed so that it will take only a few minutes of your time to complete. Your name is only collected to assist in organizing the data; it will then be deleted by your school administrator. Your name and the names of your students will not be shared with me or with anyone else.

Thank you in advance for your time.

Best regards,
Michael Arcidiacono

This research study has been approved by the George Fox University's Institutional Review Board committee. By filling out and submitting this form, you consent to having your de-identified data used in presenting the results at academic forums.

1. Your name

2. How many years of teaching experience do you have?
   Include this school year. Please answer in whole years (if you taught for part of a year, please round up)

3. How many years have you been teaching at your current school?
   Include this school year. Please answer in whole years (if you taught for part of a year, please round up)

4. Is your bachelor's degree in mathematics?
   Mark only one oval.
   - Yes
   - No (I have a bachelor's degree, but it is not in mathematics)
   - I don't have a bachelor's degree.
   - Other degree similar to bachelor's
5. If you have a master's degree, is it in mathematics?
Mark only one oval.

- Yes
- No. (I have a master's degree, but it is not in mathematics)
- I don't have a master's degree
- Other degree similar to master's

6. Please check the country where you received your teacher training/certification
Teaching certification or teaching license
Check all that apply.

- Brazil
- United States
- Other
Appendix D: Teacher demographics (figures)

Figures 1 & 2: Box plot (bachelor’s degree & master’s degree)
Appendix D: Teacher demographics (figures)

Figure 3: Box plot (country of certification)
Appendix E: Assumptions of multiple regression

Figures 4 & 5: Partial regression plot (experience & tenure)
Appendix E: Assumptions of multiple regression

Figure 6: Studentized residual plot
Appendix E: Assumptions of multiple regression

Figure 7: P-P plot