PUPIL, TEACHER, AND SCHOOL FACTORS THAT INFLUENCE STUDENT ACHIEVEMENT ON THE PRIMARY LEAVING EXAMINATION IN UGANDA: MEASURE DEVELOPMENT AND MULTILEVEL MODELING

A dissertation submitted to the Kent State University College of Education, Health, and Human Services in partial fulfillment of the requirements for the degree of Doctor of Philosophy

By

Pius Ochwo

August, 2013
Dissertation written by

Pius Ochwo
B.A., Makerere University, 2001
M.A., Uganda Martyrs University, 2006
PH.D., Kent State University, 2013

Approved by

__________________________________, Co-Advisor, Doctoral Dissertation Committee
Aryn C. Karpinski,

__________________________________, Co-Advisor, Doctoral Dissertation Committee
Tricia Niesz

__________________________________, Member Doctoral Dissertation Committee
Andrew Lepp

Accepted by

_________________________, Chair School of Foundations, Leadership
Shawn Fitzgerald and Administration

_________________________, Dean, College of Education, Health and Human Services
Daniel F. Mahony
This study examined the multilevel factors that influence mathematics and English performance on the Primary Leaving Examinations (PLEs) among primary seven pupils (i.e., equivalent to the United States [U.S.] 7th graders) in Uganda. Existing student state test data from the Wakiso District were obtained. In addition, a newly created Teacher Quality Measure (TQM) was used to collect teacher data from the same district. Pupil data from primary seven (7th grade) and the TQM data were analyzed via Rasch Analysis, Analysis of Covariance, and Hierarchical Linear Modeling to investigate the following two main objectives: (1) Developing a behavioral frequency measure of teacher quality for Ugandan teachers, (2) Examining the relationship between pupil-, teacher-, and school-level factors on pupil achievement on the PLEs in Uganda.

Specific to the first objective, it was found that a psychometrically sound measure of teacher quality can be developed. The results rendered a 38-question measure focusing on four domains: (1) Teacher Planning and Preparation, (2) Classroom Environment, (3) Teacher Instruction, and (4) Teacher Professionalism.

The second objective found that there are no significant differences between boys and girls on English achievement controlling for prior ability in English. However, there were significant differences between the sexes on mathematics achievement, with boys
having higher scores. Additionally, the results showed that there is a significant relationship between student SES (i.e., boarding and day schools) and student achievement, with higher SES students (i.e., boarding schools) having higher achievement. It was also found that teacher TQM scores were a significant predictor of student PLE mathematics and English test scores, with higher teacher quality rendering higher student mathematics and English scores. There was also a significant difference between school types (i.e., urban and rural) on student achievement in mathematics, with rural schools (i.e., lower SES schools) having higher means compared to urban schools.

Future research should continue to define the network of relationships between pupil-, teacher-, and school-level factors and pupil achievement, and maintain the measure revision and validation process of the TQM. Assessment is becoming commonplace in the classroom in Uganda, and the need to examine the influence of the teacher on pupil achievement is in high demand. Results from this study can provide insight into the disparities involving sex, student SES, and school SES that influence pupil achievement in Uganda. The findings also support administrative demands for more efficient ways to monitor teacher quality, and in turn, meet educational standards and increase student achievement.
# TABLE OF CONTENTS

List of Figures ........................................................................................................ vii
List of Tables ........................................................................................................ viii
Dedication ............................................................................................................... ix
Acknowledgments ................................................................................................ x

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>INTRODUCTION</td>
</tr>
<tr>
<td></td>
<td>The Context of the Ugandan Schooling System</td>
</tr>
<tr>
<td></td>
<td>The Context of the Study</td>
</tr>
<tr>
<td></td>
<td>Statement of the Problem</td>
</tr>
<tr>
<td></td>
<td>Research Objectives</td>
</tr>
<tr>
<td></td>
<td>Research Questions</td>
</tr>
<tr>
<td></td>
<td>The Rationale for the Study</td>
</tr>
<tr>
<td></td>
<td>Significance of the Study</td>
</tr>
<tr>
<td></td>
<td>Research Paradigm</td>
</tr>
<tr>
<td></td>
<td>Variables in the Study</td>
</tr>
<tr>
<td>2</td>
<td>LITERATURE REVIEW</td>
</tr>
<tr>
<td></td>
<td>Pupil Achievement Studies in Uganda</td>
</tr>
<tr>
<td></td>
<td>Initial Studies on Pupil Achievement in Uganda</td>
</tr>
<tr>
<td></td>
<td>Section One: Pupil-Level Variables</td>
</tr>
<tr>
<td></td>
<td>Pupil Socioeconomic Status (SES) and Achievement</td>
</tr>
<tr>
<td></td>
<td>Pupil Gender and Achievement</td>
</tr>
<tr>
<td></td>
<td>Section Two: Teacher-Level Variables</td>
</tr>
<tr>
<td></td>
<td>Teacher Factors and Student Achievement</td>
</tr>
<tr>
<td></td>
<td>Section Three: School-Level Variables</td>
</tr>
<tr>
<td></td>
<td>School Type and Student Achievement</td>
</tr>
<tr>
<td></td>
<td>Summary</td>
</tr>
<tr>
<td></td>
<td>Conclusion</td>
</tr>
<tr>
<td>3</td>
<td>METHODOLOGY</td>
</tr>
<tr>
<td></td>
<td>Research Question 1: Development of the Teacher Quality Measure (TQM)</td>
</tr>
<tr>
<td></td>
<td>Qualitative Data</td>
</tr>
<tr>
<td></td>
<td>Quantitative Data</td>
</tr>
</tbody>
</table>
Research Question 2: Predictors of Student Achievement ........................................ 33
Data Sources: Data from Pupils .................................................................................. 33
Data Sources: Data from Teachers .............................................................................. 36
Procedure .................................................................................................................. 37
Data Analysis ............................................................................................................. 39
Research Question 1 ................................................................................................... 39
Research Question 2A .................................................................................................. 39
Research Question 2B .................................................................................................. 40
Research Question 2C ................................................................................................. 44

4 RESULTS ................................................................................................................ 46
Research Question 1 ................................................................................................... 46
Literature Review (TQM) Development ...................................................................... 47
Studies Measuring Teacher Quality .......................................................................... 49
Results from Open - Ended Questionnaire ................................................................. 54
Item Writing and Content Validity ............................................................................. 55
Pilot Study of TQM (Including Revisions Made) ......................................................... 57
Pilot Study Analysis .................................................................................................... 57
Content Analysis of the 54-Item TQM ....................................................................... 61
Final TQM Analysis .................................................................................................... 64
Research Question 2A .................................................................................................. 74
Mathematics Descriptives .......................................................................................... 74
Mathematics Assumptions ......................................................................................... 75
Mathematics ANCOVA .............................................................................................. 76
English Descriptives .................................................................................................... 77
English Assumptions .................................................................................................... 78
English ANCOVA ........................................................................................................ 79
Research Question 2B .................................................................................................. 80
Mathematics Descriptives .......................................................................................... 81
Mathematics Assumptions ......................................................................................... 82
Mathematics Unconditional Model ............................................................................ 83
Mathematics Contextual (Full) Model ........................................................................ 84
English Descriptives .................................................................................................... 86
English Assumptions .................................................................................................... 86
English One-Way Random Effects ANOVA ............................................................... 87
English Unconditional Model .................................................................................... 88
English Unconditional (Full Model) .......................................................................... 89
Research Question 2C .................................................................................................. 91
Mathematics Descriptives .......................................................................................... 91
Mathematics Assumptions ......................................................................................... 92
Mathematics ANCOVA .............................................................................................. 93
English Descriptives .................................................................................................... 94
English Assumptions ................................................................................................... 94
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DISCUSSION</td>
<td>97</td>
</tr>
<tr>
<td>Research Question 1</td>
<td>97</td>
</tr>
<tr>
<td>The 38-Item TQM</td>
<td>99</td>
</tr>
<tr>
<td>Research Question 2A</td>
<td>99</td>
</tr>
<tr>
<td>Sex and Mathematics Performance</td>
<td>99</td>
</tr>
<tr>
<td>SES and Mathematics Achievement</td>
<td>101</td>
</tr>
<tr>
<td>Sex and English Performance</td>
<td>102</td>
</tr>
<tr>
<td>SES and English Performance</td>
<td>104</td>
</tr>
<tr>
<td>Research Question 2B</td>
<td>105</td>
</tr>
<tr>
<td>Research Question 2C</td>
<td>106</td>
</tr>
<tr>
<td>School SES and Mathematics Achievement</td>
<td>107</td>
</tr>
<tr>
<td>School SES and English Achievement</td>
<td>108</td>
</tr>
<tr>
<td>Implications</td>
<td>109</td>
</tr>
<tr>
<td>Research Question 2A</td>
<td>111</td>
</tr>
<tr>
<td>Research Question 2B</td>
<td>112</td>
</tr>
<tr>
<td>Research Question 2C</td>
<td>113</td>
</tr>
<tr>
<td>Research Question 2 (All)</td>
<td>114</td>
</tr>
<tr>
<td>Limitations and Future Directions</td>
<td>115</td>
</tr>
<tr>
<td>Research Question 1</td>
<td>115</td>
</tr>
<tr>
<td>Research Question 2 (All)</td>
<td>118</td>
</tr>
<tr>
<td>Conclusion</td>
<td>121</td>
</tr>
</tbody>
</table>

APPENDICES ........................................................................ 123

APPENDIX A ORIGINAL 80-ITEM TEACHER QUALITY MEASURE (TQM) .... 124
APPENDIX B FINAL 38-ITEM TEACHER QUALITY MEASURE (TQM) ... 128
APPENDIX C APPROVAL TO CONDUCT RESEARCH IN THE WAKISO SCHOOL DISTRICT ... 136
APPENDIX D THE KENT STATE UNIVERSITY INSTITUTIONAL REVIEW BOARD APPROVAL ... 138
APPENDIX E RECRUITMENT SCRIPT ........................................... 140
APPENDIX F INVITATION TO TAKE PART IN THE STUDY ........................ 142

REFERENCES ........................................................................ 144
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Structure of the Ugandan education system (Government of Uganda, Education Act, 2008)</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>The map of persons and items for the 54-item Teacher Quality Measure (TQM)</td>
<td>62</td>
</tr>
<tr>
<td>3</td>
<td>Category probabilities (i.e., probability curves) indicating the probability of a response for the 38-item Teacher Quality Measure (TQM). These curves display the likelihood of category selection (Y-axis) by the person-minus-item measure (X-axis)</td>
<td>69</td>
</tr>
<tr>
<td>4</td>
<td>The map of persons and items for the 38-item Teacher Quality Measure (TQM)</td>
<td>73</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table   Page

1  Mean English and Math Scores in Southern and Eastern African Countries ..........8
2  Independent Variables (IVs) at each Level that Influence Pupil Achievement.......11
3  Summary Statistics for the 54-Item Teacher Quality Measure (TQM) for Persons and Items ..........................................................59
4  Item Statistics: Misfit Order (54 Measured Items) ......................................60
5  Summary Statistics for the Final Teacher Quality Measure (TQM) for Persons and Items ..........................................................66
6  Summary of Category Structure (38 Measured Items) from the Teacher Quality Measure (TQM) .........................................................68
7  Item Statistics: Misfit Order (38 Measured Items) ......................................71
8  One-Way Random Effects ANOVA Model with the Mathematics Primary Leaving Examination (PLE) ........................................82
9  Unconditional Model with the Mathematics Primary Leaving Examination (PLE) ..........................................................83
10 Contextual (Full) Model with the Mathematics Primary Leaving Examination (PLE) ..........................................................85
11 One-Way Random Effects ANOVA Model with the English Primary Leaving Examination (PLE) .........................................................87
12 Unconditional Model with the English Primary Leaving Examination (PLE).......88
13 Contextual (Full) Model with the English Primary Leaving Examination (PLE) ....90
DEDICATION

This document is dedicated to my beloved parents, Mr. Daniel Makola and Mrs. Brenda Abbo, who have already preceded us into eternity. They were always there to hold me up during difficult times. Their dedication and prayers will always remain with me, and whatever I do will be for their honor. May they rest in peace. Amen.
ACKNOWLEDGEMENTS

I am deeply grateful to Dr. Aryn Karpinski, my advisor and mentor throughout my doctoral studies who served as chair of my dissertation committee. Her commitment to nurturing the diverse interests of her students was a constant inspiration for me. Dr. Karpinski helped me to discover the scholar within, and directed my dissertation and other coursework in a way that expanded my thinking about the role of a teacher in society. I always appreciated her constructive comments, and her personal support of my academic endeavors.

I am grateful for the other members of my committee, especially Dr. Tricia Niesz and Dr. Andrew Lepp who critically read my work, and provided suggestions, which were very instrumental in shaping my dissertation. Thank you also to Dr. Jacob Barkley for serving as the Graduate faculty representative for my defense. These are lessons I will carry into my future work. I thank Dr. Marietta Daulton from Walsh University who read through my work, discussed my work with me, and so improved my research as a result.

Many people in Uganda supported this study by sharing generously of their time, often on short notice, during my fieldwork. Sr. Josephine Nalubowa volunteered to be my research assistant, and shared thoughtful insights on teacher quality that helped shape my understanding of the topic in the context of Uganda. Many others also shared their insights about pupil performance and their experiences and were very insightful. They patiently answered my questions, provided me with useful documentation, and assisted my research by introducing me to their colleagues in various education organizations and
agencies. I appreciated their deep commitment to their work, and their energy and enthusiasm in the daunting task of educating Uganda’s children.

I am blessed with a wonderful religious community (i.e., The Brothers of Christian Instruction) who encouraged me and supported my work. Brother Francis Blouin strongly encouraged me to pursue a doctoral degree, and he always supported my choices. My community members at Walsh, Brothers Guy Roddy, Marcel Sylvestre, Ernest Paquet, Dennis Katusiime, Eric Gikunda Mwithiga, and Assau Mlengule cheered me on. It would have been impossible to complete this doctoral program and dissertation without their loving support.

A large number of friends and colleagues supported my efforts throughout this process. Many colleagues in the Evaluation and Measurement Program, especially Ipek Ozer, gave me good advice. I appreciated their camaraderie. A special, small group of friends (Michael, John, Monica, Sylvia, Bert, and Amber) celebrated milestones with me along the way. My religious community members prayed for me and encouraged my work. Other family members, friends, neighbors, and professional colleagues, too numerous to name here, provided constant support of many kinds.

This research and the relationships I have been privileged to develop during this work have immeasurably enriched my life and my spirit. I will always be grateful for this opportunity to have explored with others the factors that influence pupil achievement in Uganda. Thank you all, and God bless you.
CHAPTER 1

INTRODUCTION

This study examines the multilevel factors that influence mathematics and English performance on the Primary Leaving Examinations (PLEs) among primary seven pupils (i.e., equivalent to the United States [U.S.] 7th graders) in Uganda. This study uses pupil-, teacher-, and school-level variables to examine their impact on students’ performance on the PLEs. In addition, the study uses a newly developed, comprehensive, efficient, and portable measure of teacher quality. The Teacher Quality Measure (TQM) was created specifically for this study following a comprehensive review of available literature on the topic. Specifically, the TQM was assembled following Danielson's 1996 framework for teaching and learning.

The Context of the Ugandan Schooling System

The education system in Uganda comprises four levels (i.e., pre-school, primary education, post-primary education, higher education; see Figure 1 below), which are all under the supervision of the District Education Department (DED), accountable to the Ministry of Education and Sports (MoES; Namirembe, 2005). The DED trains, registers, and approves teachers, prescribes a national curriculum, provides textbooks, and selects administrators.

Primary school education is the first official phase of the schooling system and serves pupils between the ages of 6 through 12 years (i.e., pre-school is not mandatory; see Government of Uganda, 2008). The main purpose of primary school is to prepare
pupils to participate in the social, political, and economic well-being of the country, and prepare them to be global citizens (Government of Uganda, 2008; Namirembe, 2005). The primary school curriculum has therefore been designed to provide a more functional and practical education to cater to the needs of children who complete their education at the primary school level. In addition, it caters to pupils who wish, and have the means, to continue on with secondary school education (i.e., post-primary education). Primary education is universal and free but not compulsory.

At the end of the 7th year (i.e., the end of primary education), the Uganda Primary Leaving Examinations (PLEs) are taken and the results are used to determine placement at secondary school (i.e., post-primary) on a merit basis. Pupils at the end of primary education are tested on four PLE subjects: (1) English, (2) Mathematics, (3) Science, and (4) Social Studies. Other subjects are taught, but not examined in the PLEs such as Arts.
and Crafts, Local Languages, Physical Education, and Music (Namirembe, 2005). The PLEs are similar to the U.S. end-of-year exams or graduation exams (i.e., high-stakes examinations in the U.S.).

**The Context of the Study**

Learning performance (i.e., academic performance) remains a critical concern in most developing countries that have implemented the Universal Primary Education (UPE) policy, which is a plan to provide free primary education to children. Increasing learning performance, often discussed in terms of quality education, is one national strategy for economic development in Uganda. It is not only a national strategy for Uganda, but also recommended for other countries by the United Nations. That is, the 6th goal of the Dakar World Education Forum (2000) highlights the need for all countries to make efforts to “…improve all aspects of the quality of education and ensure excellence in order that recognized and measurable learning outcomes may be achieved by all, especially in literacy, numeracy and life skills” (United Nations Educational, Scientific, and Cultural Organization [UNESCO], 2000, p. 17).

Uganda implemented UPE in 1997 as part of a government plan to provide free primary education to the country’s children. Stasavage (2005) chronicled that a few months after May of 1996, the President of Uganda stated that primary school dues would be abolished for four children in each Ugandan family. The stated purpose of this was to help to raise a literate society. Soon after, the free UPE policy was implemented in 1997, the enrollment in primary education rapidly increased. Because of the policy, children from poor families started to attend schools since tuition was now provided by
the government along with their scholastic materials (e.g., books and pens). Therefore, UPE seemed successful, at least in providing tuition and support for children who formerly had not been able to attend school.

The UPE policy did not incur all positive results in Uganda (Nanyonjo, 2007). For example, one negative consequence of UPE included the massive increase of students in primary schools, which resulted in overcrowding with more than 100 pupils in a single classroom in some schools. Furthermore, while tuition in primary schools is paid for by the government, other costs such as school uniforms, bags, and lunches are not free. These extra costs limit the extremely poor children from attending school. Despite these negative consequences, implementation of UPE in 1997 has attempted to improve aspects of the quality of education in Uganda, especially in literacy and numeracy as stated by UNESCO. Of paramount importance is for researchers to begin to examine learning performance (i.e., obtaining measurable outcomes) of pupils in Uganda, which has been relatively understudied even before the implementation of UPE.

**Statement of the Problem**

The quality of an education system depends on the performance of its individual pupils, teachers whose services promote the performance of that system, and characteristics of the schools within the system. Many questions are arising from a lack of research about the performance of pupils related to the above elements of the Ugandan education system. Extremely important are concerns pertaining to whether the teachers are providing quality education and how this is reflected in the high-stakes test scores of their pupils.
Additionally, it has been demonstrated that resource allocations to UPE have increased since its inception in 1997, but with no empirical evidence of improvement in academic outcomes. Despite this increase in expenditure, it is difficult to obtain higher pupil performance on the PLEs. Therefore, it is important to conduct this study so as to examine the factors that influence pupil achievement on high-stakes tests, and thus provide evidence to inform the current educational system in Uganda.

**Research Objectives**

Researchers, educators, and administrators recognize the key role played by not only characteristics (e.g., demographic or background) of the pupils and the schools, but also the characteristics of the teachers in the success of pupils’ academic performance. However, as stated above, few studies have been conducted examining the pupil-, teacher-, and school-level influences on pupil academic performance in Uganda. Of most importance, there are no means of examining the quality of teachers in Uganda. The current study addresses this deficit through the development of a measure of teacher quality and examining the relationship between teacher quality and pupil academic outcomes on the PLE. The next section details the research questions, and subsequent sections define the specific variables used to answer the research questions.

**Research Questions**

The following research questions were used to examine the pupil-, teacher-, and school-level influences on student performance on the PLEs in mathematics and English.

1. What are the psychometric properties of a newly developed measure of teacher quality for Ugandan teachers?
2. What is the relationship between pupil-, teacher-, and school-level factors on pupil achievement on the Primary Leaving Examinations (PLEs) in Uganda?

(2A) Pupil: What is the relationship between a pupil’s sex and achievement on the PLE in Uganda?

(2B) Teacher: What is the relationship between teacher quality and a pupil’s achievement on the PLE (i.e., 7th grade) in Uganda?

(2C) School: What is the relationship between school type and a pupil’s achievement on the PLE (i.e., 7th grade) in Uganda?

The Rationale for the Study

Byamugisha (2010) reveals that despite substantial governmental reforms and expenditures on UPE and Universal Post-Primary Education and Training (UPPET; i.e., launched in 1997 and 2007, respectively), pupil underperformance remains a key issue especially in two important subjects – mathematics and English. Moreover, as mentioned previously, the enrollment of pupils at the primary level is staggering and requires urgent consideration with regards to the important influences on student academic performance. Uganda registers over 8 million children in the cycle of primary education alone (MoES, 2010). Because of the large amount of pupils that are impacted, it is essential that performance at the primary level of education be investigated, with attention given to the factors that influence their academic performance on high-stakes tests (i.e., the PLEs).

Some empirical studies in Uganda have examined the school’s environmental characteristics and how they influence student academic performance (Byamugisha & Ssenabulya 2005; Nanyonjo, 2007), with no studies specifically examining the key role
played by the teacher. Other studies have focused on pupils’ attitudes towards a subject and pupils’ socioeconomic status as a predictor of student attendance (Nanyonjo, 2007). Other factors examined in studies on the education system in Uganda include the overcrowding of classrooms leading to a lack of adequate instructional materials, and the unacceptably high pupil-to-teacher ratio that prevent teachers from offering personalized attention to students (Nanyonjo, 2007). As evidenced, there is an obvious deficit in the research base examining not only pupil- and school-level factors that influence pupil academic performance, but most notably, the impact of the teacher on pupil performance on high-stakes tests in Uganda.

In particular, performance on high-stakes tests in mathematics and English are key concerns because these are core subjects among the mandatory courses in elementary and secondary levels of education in Uganda. There is also additional time allocated to them in the school schedules. Moreover, the desired level of success in Uganda falls far below international standards. Ugandan pupils perform poorly on national examinations (e.g., the PLEs) and rate poorly internationally. In 2005, Uganda was 8th and 9th in English and mathematics among twelve countries in Africa that the South African Consortium for Monitoring Education Quality [SACMEQ] ranked (Kasirye, 2009). With national average scores of 482 in English and 506 in mathematics, Uganda ranked far below its immediate neighbors in East Africa (see Table 1).

**Significance of the Study**

According to Table 1 above, it is demonstrated that mathematics and English education in Ugandan primary schools is lower than in other countries. Byamugisha
Table 1

*Mean English and Math Scores in Southern and Eastern African Countries*

<table>
<thead>
<tr>
<th>Country</th>
<th>English Score</th>
<th>Rank</th>
<th>Math Score</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seychelles</td>
<td>582</td>
<td>1</td>
<td>554</td>
<td>3</td>
</tr>
<tr>
<td>Kenya</td>
<td>546</td>
<td>2</td>
<td>563</td>
<td>2</td>
</tr>
<tr>
<td>Tanzania</td>
<td>546</td>
<td>3</td>
<td>522</td>
<td>5</td>
</tr>
<tr>
<td>Mauritius</td>
<td>536</td>
<td>4</td>
<td>584</td>
<td>1</td>
</tr>
<tr>
<td>Swaziland</td>
<td>530</td>
<td>5</td>
<td>516</td>
<td>6</td>
</tr>
<tr>
<td>Botswana</td>
<td>521</td>
<td>6</td>
<td>513</td>
<td>7</td>
</tr>
<tr>
<td>Mozambique</td>
<td>517</td>
<td>7</td>
<td>530</td>
<td>4</td>
</tr>
<tr>
<td>South Africa</td>
<td>492</td>
<td>8</td>
<td>486</td>
<td>9</td>
</tr>
<tr>
<td>Uganda</td>
<td>482</td>
<td>9</td>
<td>506</td>
<td>8</td>
</tr>
<tr>
<td>Zanzibar</td>
<td>478</td>
<td>10</td>
<td>478</td>
<td>10</td>
</tr>
<tr>
<td>Lesotho</td>
<td>451</td>
<td>11</td>
<td>447</td>
<td>11</td>
</tr>
<tr>
<td>Namibia</td>
<td>449</td>
<td>12</td>
<td>431</td>
<td>14</td>
</tr>
</tbody>
</table>

*Note.* This is from Kasirye (2009) in the South and East African Consortium for Monitoring Education Quality (SACMEQ; 2005) report.

(2010) has noted that the lack of appropriate educational research prevents the establishment of intervention policies that could resolve the situation. Furthermore, even though some research has shown a correlation between student achievement and various teacher characteristics like qualification, in-service training, teachers’ ages, experience, tenure, and distance from school (Nanyonjo, 2007), other studies have indicated different
results depending on the country and their economic (or other environmental) situation (Byamugisha, 2010). Based on these inconsistent results and the lack of scholarly research in quality teaching and learning in Uganda, this study is unique as it will examine factors not previously examined that may impact the quality of primary education in Uganda.

While contributing to the body of knowledge on education in Uganda, this study is particularly useful to the Ugandan MoES and the proprietors involved in teaching services. The results can serve as a guide in designing interventions capable of improving the quality of teachers and how this may impact student performance. In addition, the study contributes to the accountability process which is so crucial to the stakeholders of student learning (i.e., donors, the government, parents, and the pupils themselves). Some of these interested parties have invested significant funds and, therefore, need to know why some of their resources do or do not seem to be used effectively.

Because this study investigates pupil performance within the framework of UPE, it contributes to policy and practical implementation of the UPE program. The focus on the pupil, teacher, and school (i.e., a multi-level approach), provides insight into the many influences on pupil academic performance under UPE and provide suggestions based on empirical results to improve the implementation of UPE. Therefore, the findings of this study provide a vital resource to a broad audience of educators, policy makers, and funding partners.

Finally, this study provides a timely contribution to the body of knowledge on educational reforms in Uganda, since multi-level studies on the factors that influence
pupil performance in English and mathematics on the PLEs are not available at this time. Most studies address enrollment issues, pupil-teacher ratios, and other factors that influence student performance (Nanyonjo, 2007). However, these studies have not investigated the relationship between the multi-level factors that influence pupil academic performance on high-stakes exams such as the PLEs in mathematics and English, especially the influence of the quality of the teacher.

**Research Paradigm**

As stated earlier, the purpose of this study was to examine pupil, teacher, and school factors that influence mathematics and English performance on the PLEs. This study adopted a quantitative approach, which uses Post-Positivist claims of knowledge (Creswell, 2009). For example, some of the Post-Positivist features include cause and effect thinking, reduction of variables, hypotheses and questions, use of measurement and observation, and the testing of theories (Creswell, 2009). The approach may employ survey methodology or an experimental approach to collect data. This study used a newly developed instrument to collect and to generate quantitative data on teacher quality. In addition, the national Ugandan tests in English and mathematics for pupils in primary seven (i.e., 7th grade) also provided quantitative data.

**Variables in the Study**

For the first research question (i.e., “What are the psychometric properties of a newly developed measure of teacher quality for Ugandan teachers?”), the variable is the construct of teacher quality, which was measured by the newly developed Teacher Quality Measure (TQM). This measure is explained in detail in the Methods section.
For the second research question (i.e., “What is the relationship between pupil-, teacher-, and school-level factors on pupil achievement on the Primary Leaving Examinations (PLEs) in Uganda?”) the variables are listed for Research Questions 2A through 2C below (see Table 2 below for a summary). The variables for Research Question 2A (i.e., “What is the relationship between a pupil’s sex, and achievement on the PLE (i.e., 7th grade) in Uganda?”) include:

Table 2

*Independent Variables (IVs) at each Level that Influence Pupil Achievement*

<table>
<thead>
<tr>
<th>Level</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>School</td>
<td>School Type – Urban or Rural</td>
</tr>
<tr>
<td>Teacher</td>
<td>Scores on the Teacher Quality Measure (TQM)</td>
</tr>
<tr>
<td>Pupil</td>
<td>Sex – Male or Female; Socioeconomic Status – Boarding or Day</td>
</tr>
</tbody>
</table>

*Note.* The Dependent Variable (DV) for all of these levels is pupil performance on the Primary Leaving Examinations (PLEs) in mathematics and English.

1. Pupil Mathematics and English Achievement on the PLE – The variables of pupil mathematics and English achievement was used as continuous Dependent Variables (DV$s$). The scores representing the variable of achievement were generated by the Uganda National Examinations Board, and were provided by the Wakiso School District.

2. Sex – The variable of sex of a pupil as reported by the schools to the testing agency were used (i.e., Male or Female) as the main categorical Independent Variable (IV) of interest. Sex is hypothesized to be an important variable in explaining differences at the pupil level on academic performance because
research indicates that females outscore males in tests of verbal ability (i.e., English), whereas males score higher than females in mathematics (Barrow, Boyle, Ginsburg, Leu, Pier & Price-Rom, 2007; Ganimian, 2009; Marks, 2008; Michaelowa, 2004; Mullis, 2007).

3. Prior Mathematics and English Achievement/Knowledge – The variables of pupil prior mathematics and English achievement/knowledge were used as continuous IVs that are Covariates. The scores representing the variable of prior achievement/knowledge are generated from a test (i.e., PLE practice tests or pretests) taken prior to the official PLEs. This variable controlled for student prior mathematics and English achievement/knowledge.

4. Socioeconomic Status (SES) – The variable of pupil SES as reported by the Wakiso School District were used (i.e., Day School or Boarding School) as a categorical IV that is also a statistical control (i.e., not the main IV of interest). SES was also hypothesized to be an important variable in explaining differences in pupil academic performance. Research indicates that more affluent (i.e., Boarding School) pupils outscore poorer (i.e., Day School) students on both mathematics and English exams (Tumwebaze, 2012).

The variables for Research Question 2B (i.e., “What is the relationship between teacher quality and a pupil’s achievement on the PLE (i.e., 7th grade) in Uganda?”) include:

1. Pupil Mathematics and English Achievement on the PLE (see above).
2. Teacher Quality – The variable of teacher quality was measured by the newly developed TQM (i.e., the continuous IV of interest). The variable is operationalized by a total score for teacher quality combining the following domains of the TQM: (1) teacher preparation and planning, (2) teacher classroom environment, (3) teacher instruction, and (4) teacher professionalism.

3. Prior Mathematics and English Achievement/Knowledge (see above).

The variables for Research Question 2C (i.e., “What is the relationship between school type and a pupil’s achievement on the PLE (i.e., 7th grade) in Uganda?”) include:

1. Pupil Mathematics and English Achievement on the PLE (see above).

2. School Type – The variable of School Type as reported by the Wakiso School District will be used (i.e., Urban or Rural) as the categorical IV of interest. School Type is a proxy for school SES, which was also hypothesized to be an important variable in explaining differences in pupil academic performance. Research indicates that more affluent (i.e., Urban) schools outscore poorer (i.e., Rural) school on both mathematics and English exams (Caldas & Bankston, 1997; Rumberger & Palardy, 2005)

3. Prior Mathematics and English Achievement/Knowledge (see above).
CHAPTER 2

LITERATURE REVIEW

This chapter reviews the body of literature examining the influence of pupil-, teacher-, and school-level factors on pupil achievement in mathematics and English. The organization of the literature falls into the main categories of pupil-, teacher-, and school-level factors. Each of these broader categories of literature is reviewed in relation to student achievement. Evidence from developed countries, particularly the U.S., as well as Sub-Saharan Africa was extensively used in this chapter to situate the study in a literature base.

Pupil Achievement Studies in Uganda

The National Assessment of Progress in Education (NAPE) is currently the only state research body charged with monitoring quality education in Uganda. NAPE conducts studies on the education system (i.e., the assessment of the performance of the whole education system). According to NAPE objectives (Acana, 2006), studies are conducted to determine the achievement levels of pupils on the curriculum over a period of time. The following are the latest reports of NAPE: (1) The Achievement of Primary School Pupils in Uganda in Numeracy, and (2) Literacy in English and Local Languages (NAPE, 2006).

The first NAPE study in Uganda was carried out in 1996, when the achievement of primary three (i.e., grade 3) and primary six (i.e., grade 6) pupils and their teachers
were assessed in English and mathematics (Uganda National Examinations Board [UNEB], 1997). Since then, seven more studies have been done in the same classes covering numeracy and literacy. NAPE studies (NAPE, 1999; 2003; 2005; 2006; 2007; 2008) discovered that after the introduction of UPE, there was a decline in the quality of education. The findings demonstrated pupils’ low mean scores as well as low proportions of pupils achieving the desired proficiency levels on the PLE. However, even though NAPE studies claim that the trend has generally reversed over the recent years, with more pupils reaching the required ratings, no study has been conducted to investigate the status of performance of 7th grade pupils, which is a critical year at the end of the primary level.

Another education research association in Africa is Uwezo, which means “the ability” or “capability” in Kiswahili. Uwezo is an association whose goal is to improve pupil competencies in literacy and numeracy in the three East African countries (i.e., Kenya, Tanzania and Uganda). Uwezo (2012) notes that in recent years there has been enormous growth in primary school education and this has consequently led to large increases in public budgets for education in the East African States. However, the Uwezo (2011) discovered that despite higher enrolment rates, education quality has remained low and may have declined.

For Uganda, Uwezo (2011) reported that six out of every 10 of pupils in primary three to primary seven in Uganda whose mothers had college-level education were able to read and understand a grade two level text, while seven out of every 10 of the pupils whose mothers had never acquired any level of education were not able to read and understand at a grade two level. Furthermore, the study found that pupils in lower
primary levels (i.e., grade one and two) who receive extra tuition and private tutoring were not any better in English and mathematics performance than those who did not receive extra tuition. It can be seen that Uwezo studies focused mainly on student background factors, learning inputs like books, classroom space, and have not made a thorough examination of teacher quality characteristics (e.g., teacher planning, teacher classroom environment, instruction, and teacher professionalism).

Lastly the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) seems to be the most reputable research association that has conducted more studies on pupil performance. SACMEQ draws participating researchers from 15 countries across Southern and Eastern Africa (i.e., Botswana, Kenya, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, Tanzania, Uganda, Zambia, Zanzibar and Zimbabwe). SACMEQ studies have been cantered on educational variables like school enrolment, pupil retention rate, and performance on national exams, gender ratios of students, teacher-student ratios, and access in terms of distance. What has not been done is to examine these factors specifically in Uganda and in connection with pupil achievement on high-stakes exams.

**Initial Studies on Pupil Achievement in Uganda**

Heyneman and Loxley (1982; 1983) conducted the early influential studies on the correlation between pupil, teacher, and school factors that influenced pupil achievement. The two authors examined the influence of country SES and school factors on pupil achievement in Uganda and other similar countries and found that country SES has less influence on pupil achievement than school and teacher quality. The authors argued that
“…the poorer the national setting in economic terms, the more powerful this (school and teacher quality) effect appears to be” (Heyneman & Loxley, 1983, p. 1184). This conclusion is commonly referred to as the “Heyneman–Loxley Effect” or “H–L Effect” in educational literature and is considered important because it supports the linkage between educational achievement and socioeconomic status.

**Section One: Pupil-Level Variables**

**Pupil Socioeconomic Status (SES) and Achievement**

Pupils’ SES is one critical student factor that may impact student achievement in Uganda. Byamugisha (2010) examined the instances where parents pay additional tuition so teachers may give extra lessons to the pupils. The study found that these pupils performed better than those whose parents have low SES and cannot afford the extra fee. The MoES, however, considers this practice illegal and has put a ban on such tutorials. Byamugisha (2010) further adds that some teachers have taken on private tutoring as an opportunity to make money, causing a major concern. Moreover, the excuse that the teachers normally give is the calendar days are not enough to complete the set syllabus.

Beyond Uganda, other studies have addressed student achievement as related to SES. For example, the SACMEQ found that student SES is the strongest predictor of pupil performance in reading and the second strongest in mathematics. Internationally, SES emerges in many studies (Anderson, 1991; Lee, Zuze, & Ross, 2005; Postlethwaite & Ross, 1992) presenting the relationship between pupil backgrounds and their performance. There are many factors that are proxies for SES that have been examined in relation to pupil achievement. For example, lack of transportation and distance to school
(i.e., as an indicator of lower SES) impacts student achievement (Chowdhury, 1995). Another example is pupils having lunch at school. Byamugisa (2010) found that students who had lunch had better scores in mathematics, compared to students whose parents were unable to pay lunch (i.e., an indicator of SES). Other studies outside Uganda have found similar results concerning the relationship between students’ meals and reading and mathematics performance (Etsey, 2005; Postlethwaite & Ross, 1992).

Some studies have examined SES and student achievement in terms of the available books for students at home. Greaney and Kellaghan (2008) argue that the number of books in the home correlates positively and significantly with pupil performance in mathematics. Postlethwaite and Ross (1992) hold the same view, but add that if students can borrow books from the school library, then lack of books would not be related to SES. Nanyonjo (2007), as well, found a correlation between the number of books at home and students’ performance.

**Pupil Gender and Achievement**

Discussion of how student gender may impact educational outcomes has been a crucial topic in Ugandan education services, as differences in power and status between men and women continue to be a passionate societal issue. Notably, Ugandan pupils live in a society where gender structures are strictly defined, and gender beliefs persist when they enter the school. This section identifies research where the gender gap may impact the performance in mathematics and English.

Gender concerns are not unique to Uganda. Mathematics and science disciplines are traditionally dominated by men. Even in the most progressive societies, some
maintain that girls are mathematically inferior to boys, with more pronounced differences expected during adolescence (Zuze & Leibbrandt 2011). Girls who choose careers in these fields may face challenges, even open hostility. Recent studies have suggested that expectations for students play a crucial role in how girls and boys perform. For example, Dweck (1986) noted that girls may fail to attain full academic potential if they are made to believe that it is beyond their gender or social ability.

When Uganda embraced UPE, the challenge was not only for school children, but for quality and equality at the primary level. Gender equality took center stage among other issues. Byamugisha (2011) enumerates them as follows: (1) gender equality of access, (2) retention, and (3) performance in science and mathematics. Muhwezi (2003) notes that a number of projects dealing with the above aspects were established by the Ministry of Gender, Labor, and Social Development, to formulate policies as well as monitor progress.

There was some progress in Uganda towards greater gender equality in enrollment, especially between 2000 and 2007 (Byamugisha, 2011). However, the SACMEQ noted that the enrollment trend excluded quality in learning achievement. It can be seen that studies indicate that some school resources are critical for keeping girls in school and motivating them to perform well. For example, these resources include safety measures such as fences, and sanitation like separate bathrooms for boys and girls (Byamugisha, 2011).

Additionally, Kasente, Nakanyike and Balihuta (2003) argue that social roles of girls affect their achievement. It was noted that they are often apprentices of their
mothers. Furthermore, Kasente and colleagues (2003) indicated that management of sex-related health issues in primary education does not meet the needs of female children. The authors report that schools have not been able to furnish the students with information about maturation, not to mention the provision of these essential facilities. Moreover, UNESCO (2012) noted that girls continue to participate in school activities during this time of their maturing bodies. Also, there is a lack of proper counseling services through which all students can be guided into safe livelihood. Consequently, these factors contribute to low achievement among children, particularly girls (UNESCO, 2012).

Graven (2012) contends as well that the link between gender and student support in the home has a correlation with gender-based tasks that hampers school preparation. For instance, if girls have more domestic chores at the beginning and the end of each day, they will have less time to complete homework. For instance, these might include walking long distances to collect water or firewood, cooking, cleaning, and taking care of the younger siblings and elderly family members. Consequently, they may even be forced to miss days of school in order to focus on domestic tasks.

Specifically, studies show that the low mathematics achievement is to a great extent due to social expectations. Parents assume that boys are superior to girls in mathematics, and consequently girls lose confidence. The social psychology literature on motivation has formalized the different gender-related attitudes toward mathematics that are transferred from parent and teacher to a student. According to Dweck (1986), ability in mathematics is thought to be inborn and unchangeable, and girls are made think that
they lack this innate aptitude. Boys, on the other hand, believe that skills in mathematics can be learned and improved as a result of working hard. Treatment by teachers reinforces these perceptions. Following this reasoning, student self-perceptions suggest that when boys underperform it is because they are lazy or bored with their work, but when girls do the same it is because of limited ability. Therefore, boys are encouraged to work harder whereas girls are advised to give up and face reality.

Section Two: Teacher-Level Variables

There are no studies in Uganda that address teacher quality as a factor that influences student outcomes. However, since the release of the Coleman report in 1966, and its 1990 update, many studies have examined the impact of teachers on student achievement predominantly in the U.S. The majority of these studies have concluded that teachers are one of the most critical factors that predict the variation in student test scores, preceded only by individual and family background characteristics (Goldhaber & Brewer, 1997). Therefore, this section will review literature on teacher quality characteristics that influence student achievement.

Teacher Factors and Student Achievement

Instruction refers to the ability to create a connection between the student and the course content, and the nature of tasks conducted in the teaching and learning context (Nanyonjo, 2007). Core elements of subject matter that facilitate students’ linkage to the teaching experience include multiple elements. For example, setting tasks that are appropriately challenging, assigning work that is crucial and meaningful, building a variety into content and utilizing material that arouses curiosity and is appealing to young
people (Martin, 2002a; 2003a). These elements reflect content, subject matter, and learning tasks, which are meaningful to the student. The teacher, therefore, presents the contextual dimension of the subject matter, and emphasizes relevance and utility of what is learned (Eccles, 1983; Elliot, 1997).

There is evidence indicating that the amount of time students are exposed to instruction will have a positive effect on achievement. Students who spend a large proportion of time in learning activities designed to enhance skill knowledge are likely to attain those skills more than students who spend less time in these activities (Lockheed & Komenan, 1989; Nanyonjo 2007). A critical factor also affecting achievement is the method of instruction that the teacher employs. Trigwell and Prosser (1991) have documented that rote memory strategy, meaningless expressions, terminology and symbols dominate mathematics classes, which consequently lead to lack of understanding and logical reasoning ability.

Additionally, studies show that when teachers interact closely with the learners, there are higher chances of student success (Abbott & Ryan, 2001; Battistich & Hom, 1997). Research shows that students work best when they feel that their teachers not only like them, but also like what they do during the instruction process. As Goodenow (1993) and Teven and McCroskey (1997) note, teachers can be a strong motivator in the educational achievement of the learners. Hence, students’ confidence in their teachers is a result of teacher’s warmth during instruction (Connell & Wellborn, 1991; Flink, Boggiano, & Barrett, 1990; Kontos & Wilcox-Herzog, 1997).
Past studies have shown that instruction context includes actively listening to students’ views, allowing students to have input into decisions that affect them, getting to know students, showing no favoritism but affirming all students, accepting students’ individuality and having positive but attainable expectations for students (Flink et al., 1990; Martin, 2002a; 2003a; Teven & McCroskey, 1997). For instance, these elements are a means by which the student engages with the teacher in the teaching and learning context. As well, elements of quality instruction include providing clear feedback to students, explaining things clearly and carefully, injecting variety into teaching methods, encouraging students to learn from their mistakes, clearly demonstrating to students how schoolwork is relevant or meaningful, and allowing for opportunities to catch up (Martin, 2002a; 2003a).

Finally, research has noted instructional strategies indicative of quality teaching that facilitate students’ achievement and learning experiences. For example, some strategies include setting tasks that are appropriately challenging, assigning work that is weighty and meaningful, building a variety into content and assessment tasks, and utilizing material that arouses curiosity and is fascinating to young people (Covington, 1998; Martin, 2002a; 2003a; McInerney, 2000). Studies further contend that quality teachers focus on the needs of the contextual dimension of the subject matter and emphasize the relevance of what is learned (Eccles, 1983; Elliot, 1997; McInerney, 2000; Wigfield & Tonks, 2002). Moreover, all the above research containing these concepts and strategies are consistent with Danielson’s framework of teaching (1996), which informs the measurement of teacher quality in this study.
Châu (1996) examined the importance of subject content on teacher performance. However, the author noted that the formal level of education of the teacher is not necessarily synonymous with quality or competence. Moreover, Châu’s classroom observations in different countries show that certain teachers have an insufficient mastery of the subject matter that they teach. In addition, the author found that many teachers lack the pedagogical knowledge required for good presentation of the material. Therefore, besides professional training, subject knowledge plays a crucial role in teacher performance and has an impact on the quality of teaching and the teacher.

As noted before, no studies have addressed teacher quality in relation to student achievement in Uganda specifically. However, studies completed in two African countries, namely Tanzania and South Africa, show that teacher professional development is a predictor of student academic performance, especially in critical subjects (i.e., reading and mathematics). The study found that more than three years of training had a positive impact on student performance in mathematics. Similarly, other studies in the U.S. have found that teacher professional development can impact student achievement (Hargreaves & Fullan, 1992; Raudenbush, Eamsukkawat, Di-Ilbor; Kamali, & Taoklam, 1993; Sander & Horn, 1998). As well, in a study conducted in West Africa (Greaney & Kellaghan, 2008), teachers’ initial and in-service training is shown to be vital in shaping student achievement.

**Section Three: School-Level Variables**

Past studies point to school-level factors as key in students’ achievement (Kyriakides & Luyten, 2006; Scheerens & Bosker, 1997). For example, Scheerens and
Bosker (1997) contend that an average school will explain about 19% of the variance in student academic achievement without partialling out of other variables. Kyriakides and Luyten (2006) found a consistent result that school-level predictors account for 50% of the variance in the cognitive growth of students. This section will review school factors that are related to student achievement.

**School Type and Student Achievement**

The results of the SACMEQ II study confirm what Kulpoo (1998) found that in many less developed countries, education in rural areas is often synonymous with a poor context for learning. In the latter half of the 1990s, primary school students in rural areas of Sub-Saharan Africa consistently underperformed their urban counterparts by substantial margins. Foster (1977), as cited in Zhang (2006), refer to these disparities related to school location as one of the factors that contribute towards weak student achievement. The study indicated that more achievement inequality arises more from regional disparity of the school than it does from individuals’ characteristics, such as social class and social ethnicity.

As noted previously, according to Heyneman and Loxley (1983), schools tend to play greater roles in determining pupils’ learning achievement in poor countries than in wealthy countries. The explanation given is that schools in poor countries vary more widely than those in wealthier countries in terms of their quality, in their use of trained teachers, and in materials. In poor countries, the study adds, the school makes a greater difference in how much content pupils learn than it would in a wealthier country.
Zhang (2006) examines regional disparities in that pupils living in isolated or rural areas usually have a lower SES. That is, pupils from rural areas of less developed countries often suffer a socioeconomic disadvantage. Zhang (2006) noted that, on average, students attending rural schools came from families with lower levels of SES in every system. For instance, the author contends that families of 6th-graders in Mauritius and Seychelles were, on average, much better off than those in other countries, especially Malawi, Mozambique, Uganda, and both the Tanzanian mainland and Zanzibar. However, while pupils’ families in Mauritius and Seychelles were better off, the same author found that the gaps between the well-off and not so well-off were much smaller in these countries than the case in Botswana, Namibia, and South Africa.

The same author stresses the relations in the SACMEQ study between rural students and SES. That is, rural students not only lagged behind their counterparts in reading ability but also compared unfavorably in the school conditions that are important to academic success in general. The SES levels of the families of rural students were lower, and the rural students tended to have less home support for their academic work. In addition, rural students tended to be older than their urban counterparts, as a result of late entry into the school system. Consequently, this results into a higher incidence of grade repetition.

Numerous studies have shown that school SES has a link with academic achievement. Some U.S. studies found that mean school SES and individual student SES are predictors of various student academic outcomes (Caldas & Bankston, 1997; Rumberger & Palardy, 2005). Conversely, Sirin’s (2005) meta-analysis of 74 studies
conducted between 1990 and 2000, found that school SES is more strongly correlated with academic performance than individual student SES. Moreover, the results from all three cycles of the Program for International Student Assessment (PISA) show that, in most countries, academic outcomes tend to be more strongly associated with school SES than with individual students’ socioeconomic background (Organization for Economic Co-operation and Development [OECD], 2004; 2005; 2007).

While the research literature clearly shows that school SES is a determinant of student outcomes, no findings show whether the association is consistently strong for all students. McGaw (2007) suggested that the association between school SES and academic achievement is stronger for lower SES students than for their more privileged peers. Others, such as Opdenakker and Van Damme (2001), found similar results in that SES is a critical determinant of student achievement. This is evident in the results of the study that compared ability level of students from high SES to those from low SES in Belgium. McConney and Perry (2010) found that the association between mathematics achievement and mean school SES was particularly strong for low SES students who had high levels of mathematics self-efficacy.

On the other hand, Rumberger and Palardy (2005) found that the association between school SES and student achievement was equally strong for both low and high SES students in the U.S. McConney and Perry’s (2010) secondary analysis of PISA in 2003 also found that the association in Australia is similar for all students regardless of their social background. There are other studies showing that the achievement of all students is greater in high SES schools (Lauder & Hughes, 1999; OECD, 2004; 2007),
but the link does exclude other factors that may account for variances in student achievement.

**Summary**

The purpose of this chapter has been to review studies that have examined the relationship between pupil-, teacher-, and school-level variables and student achievement. In the studies of the relationship between gender and pupil achievement, it was found that there is a range of differences in the mean scores of achievement examinations between girls and boys with boys performing better in mathematics and girls performing better in reading and English in general. Additionally, past research indicated that the teacher is one of the most critical factors in pupil achievement. As for the studies that examined the relationship between school type and student achievement, it also was found that there is a range of differences in the mean scores of achievement examinations between schools categorized as either rural and those identified as urban (i.e., low and high SES).

**Conclusion**

Even though it seems likely that the relationship between student-, teacher-, and school-level variables and student achievement is strong, it may well be the case that it is particularly strong for some countries. As such, it is likely that the relationship may vary across nations, but not enough research has been conducted to test that hypothesis. Overall, it is the goal of this study to provide an understanding of how achievement is influenced by the student-, teacher-, and school-level variables specifically for Uganda.
CHAPTER 3

METHODOLOGY

As indicated earlier, the goal of this study was twofold: (1) To examine the psychometric properties of a newly developed teacher quality measure, and (2) To identify the relationship between key student-, teacher-, and school-level factors and mathematics and English achievement among 7th grade students in Uganda. It was critical to identify the nature of these relationships to student achievement so the education stakeholders can focus policies that can improve education quality for all children in Uganda. In order to achieve the above goals, a series of analyses were used, and the data were analyzed using Rasch Analysis, ANCOVA, and multilevel modeling procedures. This chapter includes a description of the research design, setting, and population and sample of this study. It also identifies data used for this study, the method of collecting data, and the statistical methods used to analyze the data. For Research Questions 1 and 2 below, participants and materials will be discussed within each section.

Research Question 1: Development of the Teacher Quality Measure (TQM)

A measure of teacher quality was developed. The purpose of this part of the investigation was to use a two-phase exploratory mixed-methods design to develop a comprehensive, efficient, and portable measure of teacher quality to produce scores that can be used to measure the construct. The new Teacher Quality Measure (TQM) was created after a comprehensive review of the available literature on teacher quality and
exploration of Ugandan teachers’ views about teacher quality. Following this multi-
method approach, an instrument was developed (Creswell, 2009). In addition, the TQM
was constructed following the model of Danielson's framework for teaching and learning
(1996). This model was chosen because, apart from offering a general framework for
teacher evaluation, it also assesses the learning and teaching facilities that the teacher uses,
which was deemed important by Ugandan teachers.
Specifically, the aim was to develop a behavioral frequency measure of teacher quality
for upper primary schools (i.e., grades 5 to 7) in Uganda. This portion of the study
provides an in-depth description of the construct of teacher quality. Using Rasch Analysis
provided an efficient way to measure the construct using qualitative information from
interviews with Ugandan teachers, and quantitative responses (i.e., a pilot study) to the
newly developed measure. The main questions addressed for this part of the study
included the following: (1) How is the construct of teacher quality defined in Uganda?,
and (1) What are the psychometric properties of a newly developed measure of teacher
quality in Uganda?

The demand for schools to be accountable for student performance has increased
considerably in Uganda, and one way to examine if attainment of educational standards
and objectives had been actualized was through measuring teacher characteristics and
attributes such as teacher quality (Nanyonjo, 2007). The purposes for developing this
measure were multi-faceted: (1) A measure of teacher quality in Uganda did not exist,
(2) A quick and portable measure would potentially allow schools in Uganda to examine
teacher quality, diagnose teacher performance problems, and remedy weaknesses, and
(3) The measure was designed in a flexible manner so as to be used in similar countries, and be adaptable to multiple content areas such as science and social studies (i.e., besides the focus of the current study, which was mathematics and English).

**Qualitative Data**

The overall process of the development of this measure can be summarized by DeVellis (2003). A specific research objective was identified, which included identification of items that would best measure the themes presented in the qualitative results (i.e., a literature search and survey of Ugandan teachers). It should be noted that Danielson’s (1996) model was selected to frame the construction of this measure along with data from the understanding of teacher quality from the Ugandan teacher point of view. Thus, no specific item from Danielson’s framework was used.

Surveys were conducted with several primary school teachers and administrators in Uganda. Survey questions were developed from an extensive literature review on teacher quality and effectiveness. Using a data reduction technique (i.e., Qualitative Data Analysis [QDA]; Caudle, 2004), the surveys were transcribed and common themes emerged that were analogous to Danielson’s (1996) four domains. Questions were compiled under each of the above domains to create a pilot measure containing 80 items. This preliminary measure was distributed to five Ugandan teacher volunteers who were asked to review the measure for several validity criteria (Fowler, 2002). The feedback provided by the reviewers was used to improve the survey.

The volunteers included two females and three male educators with an average of 20 years of experience in their profession/field. More specifically, these reviewers
included: (1) an Associate Dean in the School of Education (i.e., 30 years of experience), (2) an Associate Registrar (i.e., 20 years of experience), (3) a teacher of English (i.e., 8 years of experience), (4) a retired principal (i.e., 40 years of experience), and (5) a teacher of mathematics (i.e., 2 years of experience).

The revised survey (see Appendix A) was administered to teachers in the Wakiso School District in Uganda. This district was selected because of the available variety of schools such as rural and urban, varying class sizes, and children of diverse backgrounds. A research coordinator asked the teachers to fill out the TQM. Participants had approximately one month to respond, and the survey took approximately 20 minutes to complete. Twenty teachers responded, with 11 (55%) males, and all professionally trained with either a bachelor’s degree or higher. All teachers taught English, and were in urban (n = 8; 40%) or rural (n = 12; 60%) schools.

Quantitative Data

Rasch Analysis was used to examine the psychometric properties of the newly developed TQM (i.e., in this piloting phase). Responses to the items in the TQM were ordered categories (i.e., a Likert scale), indicating increasing levels of a response on the variable of interest from Few Occasions (i.e., coded 0) to Almost All Occasions (i.e., coded 3). Based on the uniform response scale used across all items, a Rating Scale Model (RSM) was implemented (Andrich, 1978). The Rasch Analysis of the TQM produced fit statistics, infit and outfit, which were examined to determine items or persons that were problematic for model fit. Based on this information, items or persons were eliminated to produce the best possible model, and the remaining items comprised
the revised 38-item TQM (see Appendix B). Thus, teacher scores could range from 0 to 114, with higher scores representing a higher reported frequency of behaviors indicative of more teacher quality. The results from the development of the TQM including the Rasch Analysis are presented in chapter four of this document.

**Research Question 2: Predictors of Student Achievement**

In this portion of the study, existing and collected data were used to examine the factors that impact student achievement in mathematics and English (i.e., pupil, teacher, and school factors).

**Data Sources: Data from Pupils**

Data were from the Uganda National Examinations Board (UNEB) from the PLEs in the current academic year. This was approved by the Wakiso School District (see Appendix C). The PLE is a nationally standardized assessment that measures students’ capabilities in mathematics, English, social studies, and science. According to Acana (2006), the PLE is a tool for placement into the next level of education (i.e., secondary level) as well as providing information on students’ achievement at the end of the seven-year program (i.e., the primary level). Acana adds that UNEB exams focus on young students’ abilities to use their knowledge and higher-order thinking skills.

At the end of every academic year, in the beginning of November in Uganda, the primary seven pupils take the PLEs to mark the end of the primary level of education. The best possible test score a pupils can attain is a total of four (i.e., implying that a pupil earns one point in each subject). The worst grade is a total of 36 (i.e., nine points for each subject, which is failing). Pupils with grades ranging from four to 12 points are
considered to have attained a first class grade. Those with scores of 13 to 23 get a second class grade; 24 to 29 get a third class grade; and those with 30 to 34 pass with a fourth class grade. That is, the smaller the grade number, the better the result. A candidate cannot attain a grade higher than a score of 4 because they have to take tests in four subjects.

As mentioned above, the PLE results are categorized into six score levels that are known as “Divisions.” Divisions I through IV are for those the pupils who are considered to have attained passes. Conversely, Divisions U and X are for the pupils that have not passed. Pupils, who score 4 or 5, are considered the best within the very highest division. This means that they may be accepted by the most prestigious secondary schools. The MoES carries out a selection process to allocate the successful candidates to the schools of their choices, if they meet the schools' cut-off score. Therefore, these tests are considered very high-stakes and analogous to the ACT or SAT in the U.S.

The PLEs are scheduled for two days. Mathematics and social studies are done on day one, whereas science and English are on day two. The English exam has two key sections. The first section is comprised of fifty questions. The questions are constructed around topics like alphabetical-ordering knowledge, finding the correct form of words, organizing of words in sentences, abbreviations, and singular versus plural of irregulars (e.g., sheep versus sheep). The second section is reading comprehension and is comprised of five different passages, which are read by the student. After reading the passages, a student has to complete a series of questions about each passage. Similarly, the mathematics exam has two sections. The first section consists of 30 questions of short
and straightforward math problems such as arithmetic, percentages/ fractions, geometry, and roman numerals. The second section consists of 12 questions of complex concepts like creating accurate angles (i.e., geometry).

Beginning in 1981, the PLE was administrated every year to groups of 12-year-old students across Uganda. In this proposed study, the 2012 mathematics and English exams will be utilized. The population includes approximately 1,000 7th grade students from 20 schools. The argument for focusing on primary seven is that it marks the end of the primary course cycle, and this includes the mass implementation of a national, standardized, high-stakes test (i.e., the PLEs). Also, because this is a national exam with a high number of examinees, it offers enough data to evaluate teaching and learning in a multi-level manner. Moreover, even though there are several alternative forms of basic education in Uganda, primary education is overwhelmingly the main segment of education.

Data from pupil practice test were used as well. In Uganda, the primary seven practice exam is done prior to the PLE. The practice test is popularly referred to as "mock exam." It is similar to a “practice run,” which is prepared by a team of teachers at the district level. The mock exam is a trial exam as the questions and the scope are similar to the final PLE exam. Moreover, the mock exam is given to the registered candidates under the same conditions as the actual PLE exam. The idea is to prepare the pupils for high-stakes test conditions and discipline. The mock exam provides teachers as well as the pupils with feedback, which allows individuals to gauge their relative strengths and weaknesses in the subject content.
Data Source: Data from Teachers

Thirty-six teachers were surveyed (i.e., using the TQM) from the Wakiso School District in Uganda. This district was selected because it offers a variety of schools such as rural and urban, highly and poorly performing schools, schools with teachers of varying qualifications, children of various economic, social, and academic backgrounds, and boarding and day schools. Of the 36 teachers surveyed, a larger percentage was male. This is not surprising because Uganda’s upper primary section (i.e., 5th - 7th grades) is dominated by male teachers. All teachers forming the sample of this study were professionally trained ranging from a large number of bachelors degree holders, to general diploma holders, and fewer 8th Grade III certificate holders.

Eighteen teachers who instruct English and 18 teachers who instruct mathematics were selected. English and mathematics teachers were used because the main purpose of basic education in Uganda is the achievement of literacy and numeracy. In addition, a larger number of teachers are Math and English teachers, and a large sample size was needed for multi-level analysis. Finally, the importance that is attached to English and mathematics in Uganda’s education system was noticed from the fact that both subjects, unlike others, appear daily on the class schedule.

Diversity was considered and teachers were selected from rural and urban settings. Teachers in major content areas were surveyed in mathematics and English, and those teaching other subjects were excluded (e.g., music, art, religion, science, social studies). The reason for exclusion of non-mathematics and English teachers was that English and mathematics are considered an essential knowledge base in Uganda, and other subjects are
not the main focus of the PLEs. The study also excludes teachers from higher levels of education because the importance attached to basic education in Uganda is great. In addition, the measure created (i.e., the TQM) focuses on teacher quality in and around the 7th grade level, not older and younger grade levels.

**Procedure**

Institutional Review Board (IRB) approval was obtained from Kent State University (see Appendix D). As mentioned previously, official permission was obtained from the Wakiso School District education office to use their existing data from schools (i.e., the PLE scores; see Appendix C) and to survey their teachers (i.e., the TQM). The primary investigator contacted the administrators at each school in the Wakiso School District and asked for willingness to participate (i.e., both from the administrators and from the 7th grade mathematics and English teachers). After the schools were contacted and verbal consent to participate was achieved, the primary investigator contacted the research coordinator to administer the surveys to the identified schools in the Wakiso school district (i.e., to the teachers). The above mentioned research coordinator is a university graduate in Education and an expert in teacher training in Uganda as well as the assistant academic registrar at the Kisubi Brothers University College, a Constituent College of Uganda Martyrs University. This coordinator also completed the CITI training required by Kent State University IRB.

The research coordinator went to each school that had agreed to participate in the district to administer the surveys. She introduced herself and the study (see Appendix E), and invited participation verbally and also with an informational letter (see Appendix F).
The teachers had approximately 15 to 20 minutes to complete the survey. After completion, the surveys were mailed to the research coordinator who compiled the data and entered the information into an Excel spreadsheet. Only ID numbers were used for each teacher. After completion of the data set, it was sent to the primary investigator for analysis. This also included the compilation of existing data from the Wakiso School District (i.e., pupil PLE scores, sex, SES, pretest scores, and school type).

This study was carried out in the Wakiso school district, located in central Uganda. The district surrounds Kampala, the capital of Uganda, and boarders Mukono, in the East, Mubende and Mpigi districts in the West, Luwero district in the North, and Kalangala in the South. Formerly part of the Mpigi district, the three counties of the Mpigi district (i.e., Busiro, Kyaddondo, and the Entebbe Municipality) became the Wakiso district. The sub-counties in Wakiso are Nsangi, Wakiso, Namayumba, Kakiri, Kasanja, Masuliita, Ssisa, Katabi, Nangabo, Gombe, Nabweru, Makingdye, and Busukuma. The town councils include: Nansana, Kira, Wakiso, and Kakiri. The district has a total of 567 primary schools with 246 governments, 257 private, and 64 community schools. For secondary schools, the district has over 139 schools, with 18 being government, 73 private, and 48 community. It also has one teacher-training college, and two technical institutes (i.e., Nkumba University and Fisheries Training Institute).

The Wakiso district was selected because of its distinguishing features. These included geographic and rural/urban divergence as well as its educational achievement disparities. For example, even though the district boasts some of the best performing schools in the country, it has also some of the most under-performing schools. Moreover,
despite its diverse urban setting and closeness to the capital city, Wakiso also has a large rural farming area. Like elsewhere in the countryside, the main economic activity in the Wakiso villages is subsistence agriculture. These villagers still use rudimentary tools such as hand-held hoes, forks, and so on for agriculture. Many of the villages do not have electricity, tarmac, or permanent roads.

Data Analysis

Research Question 1

The first research question’s data were analyzed qualitatively and quantitatively. The data analysis procedure explained above (i.e., literature review, interviews, transcription, finding common themes, item writing, content and construct validation, revisions, piloting, Rasch Analysis) is described in the Results section. Data were analyzed according to common procedures used in developing measures to ensure validity and reliability (Crocker & Algina, 2008).

Research Question 2A

This research question (i.e., “What is the relationship between a pupil’s sex and achievement on the PLE (i.e., 7th grade) in Uganda?”) was analyzed using a Two-Factor Analysis of Covariance (ANCOVA). This was conducted to determine the relationship between a pupil’s sex and achievement on the PLE in Uganda, while controlling for pupils’ prior mathematics and English achievement/knowledge. That is, there were two IVs (i.e., the two factors) – sex and SES. There was one covariate – prior achievement/knowledge (i.e., PLE pretest scores in mathematics and English). Finally, there were two DVs – mathematics and English scores on the PLE.
Outliers were screened and assumptions were examined in the ANCOVA model including: (1) Independence, (2) Homogeneity of Variance, (3) Normality, (4) Linearity, (5) Independence of the Covariate and the IVs, and (6) Homogeneity of Regression Slopes. Outliers and assumptions were examined before generating the ANCOVA results (Lomax & Hahs-Vaughn, 2012).

As noted above, there were two DVs. Thus, two separate ANCOVAs were examined as it was not expected that there would be a strong correlation between English and mathematics scores to conduct a MANCOVA. For English, the main null hypothesis included that there was no statistically significant difference in English achievement on the PLE (i.e., after controlling for prior achievement/knowledge) between males and females. The alternative hypothesis included that there is a statistically significant difference in English achievement on the PLE (i.e., after controlling for prior achievement/knowledge) between males and females, with females having higher scores. For mathematics, the main null hypothesis included that there is no statistically significant difference in mathematics achievement on the PLE (i.e., after controlling for prior achievement/knowledge) between males and females. The alternative hypothesis included that there is a statistically significant difference in mathematics achievement on the PLE (i.e., after controlling for prior achievement/knowledge) between males and females, with males having higher scores.

**Research Question 2B**

This research question (i.e., “What is the relationship between teacher quality and a pupil’s achievement on the PLE (i.e., 7th grade) in Uganda?”) was analyzed using a
Two-Level Hierarchical Linear Model (HLM). HLM can be used to investigate the relationship between teacher quality, and pupils’ achievement on PLE achievement in mathematics and English controlling for prior achievement/knowledge. Student scores in English and mathematics on the PLE, as well as the survey scores from the newly developed TQM were used. HLM (i.e., Multilevel Modeling) was used because it is the most appropriate statistical technique for the hierarchical structure of these data such as students nested within teachers. There was one main IV or predictor in the model – teacher TQM score. There was one covariate – prior achievement/knowledge (i.e., PLE pretest scores in mathematics and English). Finally, there was one DV – mathematics and English scores on the PLE.

As mentioned above, HLM is preferred because the pupils are nested within the teachers (Hox, 2002). Because of these grouping effects (i.e., students nested within teachers), the interaction between pupils makes them, in the same group, more alike than the pupils in different groups. Consequently, the observation of pupils within each group can no longer be considered statistically independent, which means the traditional statistical approaches, like regression or ANOVA, are seriously flawed and not applicable (Raudenbush & Bryk, 2002). Furthermore, as Raudenbush and Bryk (2002) note, failure to recognize the hierarchical nature of data in educational settings, or any setting for that matter, results in unreliable estimation of the effectiveness of teachers and their practices, which can lead to misinformed educational policies or practices.

With HLMs, each of the levels in this structure was represented by its own sub-model. These sub-models express relationships among variables within a given level, and
specify how variables at a higher level influence characteristics and processes at a lower level. The HLM modeling procedure had three steps. At the first step, the analysis produced the null model (i.e., the One-Way Random Effects ANOVA) with only the pupil-level outcome variable (i.e., mean PLE achievement), and no IVs at the pupil level or teacher level. At the second step, a pupil-level model was developed without variables at the teacher level (i.e., the Random Coefficients Model also called the Unconditional Model). This step includes the effects of prior achievement/knowledge on the DVs (i.e., PLE scores). Teacher variables were added to the pupil model at the third step (i.e., the Contextual Model). This “Full” Model was created to examine how teacher quality influences pupil mathematics and English performance on the PLE controlling for prior achievement/knowledge. This is the common model-building procedure used in HLM, although the third step (i.e., the Contextual Model) is the one that specifically addresses Research Question 2B.

As mentioned earlier, Research Question 2B examined the relationship between teacher quality and a pupil achievement on the PLE. In an HLM setting, this means that the teacher-specific intercepts, which are treated as random at the second level, should be adjusted. As Raudenbush and Bryk (2002) argued, when the main interest is to “…estimate the association between a level 2 predictor and the mean of Y, adjusting for one or more level-1 covariates” (p. 142), then grand-mean centering is more appropriate. HLM is based on a linear regression model, and accordingly, the intercept is the value of Y when X is equal to zero. For many variables, a zero value is not possible. Therefore,
the continuous variable of TQM score was centered around the grand mean for an accurate interpretation.

The null hypothesis was that TQM scores will not be a significant predictor of PLE scores in mathematics and English controlling for prior achievement/knowledge. The alternative hypothesis was that TQM scores would be a significant predictor of PLE scores in mathematics and English controlling for prior achievement/knowledge, with higher TQM scores predicting higher PLE scores (i.e., a positive relationship). Two Two-Level HLMs were used. That is, estimates were derived for mathematics as well as for English in separate model-building procedures. The level 1 model included:

\[ Y_{ij} = \beta_{0j} + \beta_{1j} (\text{PRETEST})_{ij} + r_{ij} \]

where \( Y_{ij} \) is the pupil’s PLE score in either mathematics or English depending on the model being run for pupil \( i \) for teacher \( j \), \((\text{PRETEST})_{ij}\) is the control for a pupil’s pretest score in either mathematics or English depending on the model being run, \( \beta_{0j} \) (i.e., the intercept) is a pupil’s initial PLE score (i.e., in either mathematics or English), \( \beta_{1j} \) (i.e., the slope) represents the pupil’s expected change in PLE score for a one-unit increase in PRETEST, and \( r_{ij} \) are independent and normally distributed errors with a common variance (Raudenbush & Bryk, 2002).

The model at level 2 describes how the above formula tends to vary based on teacher TQM scores. The pupil-level parameters become the outcome variables in the level 2 models, where they were assumed to vary across pupils depending on teacher TQM scores. The model at Level 2 included the following:
\[ \beta_{0j} = \gamma_{00} + \gamma_{01} (TQM)_{1j} + u_{0j} \]
\[ \beta_{1j} = \gamma_{10} + \gamma_{11} (TQM)_{1j} + u_{1j} \]

where \( \beta_{0j} \) and \( \beta_{1j} \) are the pupil- and teacher-specific PLE score parameters, \( \gamma_{00} \) is the baseline expectation (i.e., initial PLE status), and \( \gamma_{10} \) is the expected linear change of PLE scores. Thus, the above formulas model how the pupils intercept and slope changes depending on teacher TQM score. Finally, \( u_{0j} \) and \( u_{1j} \) are residuals.

Research Question 2C

This research question (i.e., “What is the relationship between school type and a pupil’s achievement on the PLE (i.e., 7th grade) in Uganda?) was analyzed using a One-Factor Analysis of Covariance (ANCOVA). This was conducted to determine the relationship between school type (i.e., essentially school SES) and achievement on the PLE in Uganda, while controlling for pupils’ prior mathematics and English achievement/knowledge. That is, there was one IV – school type (i.e., urban or rural). Rural schools are typically lower SES schools, and urban schools are higher SES schools. There was one covariate – prior achievement/knowledge (i.e., PLE pretest scores in mathematics and English). Finally, there were two DVs – mathematics and English scores on the PLE.

Outliers were screened and assumptions were examined in the ANCOVA model including: (1) Independence, (2) Homogeneity of Variance, (3) Normality, (4) Linearity, (5) Independence of the Covariate and the IVs, and (6) Homogeneity of Regression Slopes. Outliers and assumptions will be examined before generating the ANCOVA results (Lomax & Hahs-Vaughn, 2012).
For English, the main null hypothesis included that there is no statistically significant difference in English achievement on the PLE (i.e., after controlling for prior achievement/knowledge) between urban and rural. The alternative hypothesis included that there is a statistically significant difference in English achievement on the PLE (i.e., after controlling for prior achievement/knowledge) between urban and rural schools, with urban schools having higher scores. For mathematics, the main null hypothesis included that there is no statistically significant difference in mathematics achievement on the PLE (i.e., after controlling for prior achievement/knowledge) between urban and rural schools. The alternative hypothesis included that there is a statistically significant difference in mathematics achievement on the PLE (i.e., after controlling for prior achievement/knowledge) between urban and rural schools, with urban schools having higher scores.
CHAPTER 4

RESULTS

In this chapter, the results addressing the research questions are presented. For Research Question 1, a combination of qualitative and quantitative methods was employed to develop and analyze the TQM. To answer Research Question 2 (i.e., A, B, and C), Analysis of Covariance (ANCOVA) as well as Hierarchical Linear Modeling (HLM) were used.

Research Question 1

Research Question 1 asked, “What are the psychometric properties of a newly developed measure of teacher quality for Ugandan teachers?” This research question was analyzed using a combination of qualitative and quantitative methods. First, the development of the TQM will be described (i.e., the qualitative component) followed by Rasch Analysis (i.e., the quantitative component) to further examine the psychometric properties. Croker and Algina (2008) do recommend that the item developer carries out specific activities in improve upon the items that represent the construct. Therefore, to refine the teacher quality construct that is being measured in this study, the researcher carried out the following activities: (1) review of research, (2) content analysis, and expert judgment. Particularly, the analysis approach used in this study is drawn from Croker and Algina (2008), and it is why the literature review for TQM is presented in this section.
Literature Review (TQM Development)

Defining teacher quality, although no simple task due to its subjectivity, includes ample teaching experience (Murnane & Phillips, 1981), teachers who have strong academic skills (Ehrenberg & Brewer, 1994; Ferguson & Ladd, 1996; Hanushek, 1996), and the appropriate formal training in the teacher’s main content area (Ingersoll, 1999). These indicators of quality are components in the definition of highly qualified teachers in the No Child Left Behind (NCLB) Act (NCLB, 2001). Other indicators of quality teaching have been outlined in Danielson’s (1996) four domains: (1) planning and preparation, (2) classroom environment, (3) instruction, and (4) professional responsibilities. Charlotte Danielson’s Framework of Teaching has been used as a means for measuring the effects of teaching practice through classroom observation (Strong, 2011).

The Instructional Quality Assessment (IQA) toolkit has been used to measure teacher quality. The IQA toolkit takes into consideration analysis of teacher assignments and also pupil work samples (Strong, 2011). Also, there is the Protocol for Language Arts Teaching Observation (PLATO), which was developed for middle and high school English language arts students specifically (Grossman, 2010). PLATO was constructed considering the existing literature on effective instruction in secondary-level English. The items on this measure are intended to measure distinct aspects of classroom instruction.

Some research has attempted to develop measures of teacher quality, or also labeled as teacher performance or effectiveness. It has been challenging to quantify and measure the construct, and researchers have used several different methods including
observational evaluation rubrics, portfolios, and value-added calculations for student achievement (e.g., Raudenbush, 2004; Schacter & Thum, 2004; Tucker, Stronge, Gareis, & Beers, 2003). There are several large-scale teaching surveys that focus on measuring reform-oriented practices or enactment of curriculum from the National Center for Education Statistics (NCES), the Trends in International Mathematics and Science Study (TIMSS), and the Surveys of Enacted Curriculum (SEC). As expected, there are concerns about using self-report surveys of teacher performance (Burstein, McDonnell, Van Winkle, Ormseth, Mirocha, & Guiton, 1995), but some research has shown a high correlation between these surveys and observation (Mayer, 1999). Moreover, Strong (2011) notes that surveys may consume a large amount of time and might be less reliable.

Reviewed literature reveals that there is no evidence of a measure of teacher quality in Uganda. The most popular way of measuring teacher quality in Uganda is through academic achievement, and is usually determined by the scores from the state examination (i.e., PLEs). The higher the scores, the more it is assumed that the teacher is effective, and a greater excellence is attributed to the school (Owens, 2005). The Uganda National Examination Board (UNEB) is mandated to measure academic achievement in Uganda. The PLEs can be considered one “measure” of teacher quality in Uganda as it monitors and evaluates student academic achievement. Acana (2005), however, observes that the PLEs are intended for the selection of kids to the Secondary School Level, and other Post-Primary Vocational Institutions. A number of studies on student achievement indicate that the PLEs cannot be the best measure of teacher effectiveness because it is not directly measuring the construct (Odada 2005; Olema, 2005). Therefore, there is a
need to examine other ways of measuring teacher quality so that it can be examined in relation to pupil achievement.

Other than the PLEs, two other types of assessments have been used by the Uganda National Examinations Board (UNEB) in conjunction with the Ministry of Education and Sports (MoES) at this primary level. The first is the National Assessment of Progress in Education (NAPE), and the second is the Southern and Eastern African Consortium for Measuring Education Quality (SACMEQ). These assessments provide information that can be used as a review for quality enhancement, of which Ugandan teachers are a part.

**Studies Measuring Teacher Quality**

Schacter and Thum (2004) conducted a study that used rubrics to measure teacher quality, and concluded that it was highly predictive of pupil academic progress across elementary grades. Their sample consisted of 52 teachers from five Arizona Elementary schools. Each teacher was observed eight times over one school year. Pre- and post-scores on the Stanford 9 Achievement Test in Reading, Language, and Mathematics (i.e., the total of all) were used to measure pupil achievement. Through multilevel modeling, controlling for student demographics, the teacher quality rubric was related to pupils’ gains on the Stanford 9 Achievement Tests. Specifically, one standard deviation increase in teacher quality was related to classroom achievement gains of 6.80 points in English, 7.06 points in mathematics, and 4.73 points in reading.

Other related studies include Cohen (1994), which measured teacher’s use of instructional practices in mathematics from data collected in a 14-item survey from
elementary teachers \((N = 975)\) in California. Another study was the third TIMSS (Mullis, Martin, Smith, Garden, Gregory, Gonzalez, & O’Connor, 2003). Among other things, this study measured classroom processes by collecting perceptions of teacher strategies and teacher practices that are used in aiding learning. The results of these studies showed a small correlation between these practices and pupil achievement.

Harris and Sass (2008) interviewed 30 Florida principals about their elementary, middle, and high school teachers, asking them for overall ratings of effectiveness, and ratings on the following characteristics: caring, communication skills, enthusiasm, intelligence, knowledge of subject, strong teaching skills, motivation, works well with grade team/department, works well with me (i.e., the principal), contributes to school activities beyond the classroom, and contributes to overall school community. The results showed a statistically significant and positive correlation between principals’ overall ratings and value-added achievement scores. Although there was a relatively high correlation between the principals’ estimation of a teachers’ ability to raise pupil achievement in mathematics and the actual mathematics value-added scores, the relationship between the equivalent question and reading scores was not significant. The researchers concluded that, although principal evaluations are better predictors of pupils’ achievement than other indicators like teacher education and experience, they are not strong enough to form the basis for rating high-stakes decisions.

Wilkerson, Manatt, Rogers, and Maughan (2000) compared principal evaluations with other methods of teacher assessment for the prediction of student gains in mathematics, reading, and English. Their sample was comprised of 35 Kindergarten
through 12th grade teachers and four principals in one school district that had adopted a new evaluation process. Data were collected from principals, subordinates, and peers. In this case, compared to the previous paragraph, the results showed that the principal evaluations were not a good predictor of student learning. In another study, Murnane (1975) studied elementary teachers using a cross-sectional analysis. Murnane found that teacher evaluations were predictors of student achievement, but only to a modest degree. Moreover, Strong (2011) notes that cross-sectional analyses of student achievement have the disadvantage of not accounting for the effects of other educational inputs.

In a comparison study of 36 teachers, 18 of whom were certified in mathematics while 18 were certified in another field, Hawk, Coble, and Swanson (1985) concluded that students who were taught by in-field certified math teachers achieve at a higher level than do students taught by out-of-field teachers. Teacher differences were measured by: (1) student achievement, (2) teacher knowledge of subject, and (3) teacher professional skills as observed in the classroom. Students were given a pretest using the Stanford Achievement Test (i.e., general math) and the Stanford Test of Academic Skills (i.e., algebra). Pretest scores were not significantly different for students taught by in-field versus out-of-field teachers. However, after receiving five months of mathematics instructions, the 826 students in the study were given the same Stanford Tests as post-tests. Results of the study indicated that student achievement is greater in general mathematics and algebra when the students are taught by teachers certified in mathematics. The results of this study lend support to maintaining certification requirements as a mechanism to assure the public has qualified classroom teachers, at
least in mathematics. Likewise, Hawkins, Stancavage, and Dossey (1998) found that students, whose teachers had certification in mathematics, also had higher grades.

There is a body of work emphasizing that effective teachers express some degree of intentionality in interactions with individual children and the classroom as a whole. This attribute is manifest in a combination of high expectations, skills of management and planning, a learning orientation in the classroom, engaging children's attention with appropriate activities, and using effective, evaluative feedback and questioning, among other indicators (e.g., Berliner, 1987; Brophy & Good, 1986; Stipek, 1988). For example, Bogner, Raphael, and Pressley (2002) examined how teachers motivate students during literacy instruction in seven first-grade classrooms and found that literacy engagement was higher in classrooms in which teachers used scaffolding, encouraged autonomy and choice, had a caring manner, interacted positively with individual students, made deep and personal connections to students, encouraged creativity, and generally set a positive tone in the classroom. Similarly, Dolezal, Welsh, Pressley, and Vincent (2003) found that teachers in highly engaging third-grade classrooms provided tasks that were cognitively challenging, presented probing questions, and provided a positive classroom atmosphere. Thus, the integration of active, intentionally instructive behaviors with a socially warm and responsive interpersonal approach was closely linked to higher student performance and appears to be a hallmark of high-quality teaching.

Although the importance of teacher input is well documented in the literature, one disagreement surrounds what specific teacher attributes indicate quality, and how best to invest resources to provide that quality to benefit all students. Studies using value-
added achievement data yield indeterminate findings about the relationship between
teachers’ degree level (e.g., bachelors, masters) and coursework and observed student
achievement differences (Hanushek, 1992; Murnane & Philips, 1981). Eberts and Stone
(1984) find that the number of mathematics-related college courses taken by teachers
within the past three years is uncorrelated with elementary mathematics achievement.
Summers and Wolfe (1977) find a negative relationship between scores from the National
Teachers Examination and student achievement, while Goldhaber and Brewer (2000) do
not find the predicted influences of subject-specific teacher certification variables (i.e.,
certification types in science, history and English) on student achievement gains. Other
studies yield some positive results that are problematic due to the absence of controls for
socioeconomic status and pretest scores (Fetler, 1999).

A review of prior empirical studies on teacher qualifications by Wayne and
Youngs (2003) also reveal a mixed bag of findings. In cases of degrees, course work and
certification, the results of linking teacher credentials to improved student outcomes have
been tenuous except in teachers’ standard certification in mathematics (i.e., subject matter
coursework and passage of licensure tests). While a number of studies confirm that
children and adolescents learn more from teachers with certain characteristics, findings
from the teacher quality literature imply that strengthening teacher credentialing
requirements may enhance the qualifications and capacities teachers bring to their work
(Darling-Hammond & Youngs, 2002). However, the judicious use of this research
evidence must regard the linkage between specific teacher attributes and improved
educational outcomes as influenced by the many elements of teacher quality. Most of the
research does not seek to capture interactions among the multiple dimensions of teacher quality, and, as a result, there are significant gaps in the research that still need to be explored.

As mentioned earlier, some empirical studies in Uganda have examined the school’s environmental characteristics and how they influence student academic performance (Byamugisha & Ssenabulya 2005; Nanyonjo, 2007). However, no studies in Uganda have specifically examined the key role played by the teacher. Because of this deficit, it was necessary to examine teacher quality through the development of the TQM.

**Results from Open - Ended Questionnaire**

After examining the literature on teacher quality, a preliminary, informal investigation (i.e., interviews with Ugandan teachers) was completed. The main interview question was, “What is your understanding of a quality teacher?” This open-ended question was developed so as to get the participants perspective of a quality teacher within the Uganda context. The question was used as a flexible guide to elicit the understanding of teacher quality from the respondents (i.e., several Ugandan teachers not participating in the final study). The respondents were given two weeks to answer the open-ended question, and turned in their written responses to the primary researcher.

The sample population (i.e., using purposive sampling; Miles & Huberman, 1994) for this phase of the study was sought using the following criteria to locate experts: (1) at least 10 years of teaching experience, (2) if they have ever taught in schools that are considered to have “quality” teachers, considering their pupils scores on the PLE, and (3) when employed as teachers, if they either taught English or
mathematics. Experts were sought through convenience sampling techniques, such as invitation sent directly via email, and word-of-mouth through a research assistant. The researcher surveyed a total of 34 participants for this phase (i.e., 21 females and 13 males). Thirty-two participants were between the ages of 40 and 55, whereas two participants were between the ages of 30 and 35. Of these participants, there were 31 that identified themselves as Ugandans, one Kenyan, two Tanzanians, and one expatriate teacher, originally from Canada.

Qualitative Data Analysis (QDA) was used to examine the meaningful and symbolic content of the written responses from the teachers (Caudle, 2004). Caudle’s framework for QDA was used: (1) Data reduction and pattern identification, and (2) Producing objective analytic conclusions and communicating those conclusions. That is, data from the teachers’ written responses to the interview question were reduced to common themes. The themes that emerged were found to be consistent with the four domains of Danielson’s Teaching Framework: (1) planning and preparation, (2) the classroom environment, (3) instruction, and (4) professionalism. Items were then constructed around these main themes/domains.

**Item Writing and Content Validity**

An 80-item measure was created (i.e., the initial pool of items for the TQM; see Appendix A). The survey was a behavioral frequency format, with items written about general teacher planning and preparation, classroom environment, instruction, and teacher professionalism as indicated above. In the initial measure, participants were to be asked to rate the items on a 5-point Likert scale (i.e., 0 = Never, 1 = Few, 2 = Many, 3 =
Most, 4= Almost all). The 80-item measure was estimated to take approximately 20 to 25 minutes to complete.

Content validity was provided through volunteers who reviewed the 80-item survey (i.e., content-matter experts). First, five teachers were asked to read through the written items, to describe whether the question was clear to them, and also, to provide an assessment of how well the question addressed teacher quality. Following this process, questions were revised based on clarity and ability to match the question with the teachers’ views of a “quality teacher.” Additionally, two of the five content-matter experts (i.e., all involved in teacher education at the college level) were provided with the survey and asked to provide feedback on whether the questions covered an adequate range of teacher quality domains, whether the questions were accessing the construct of teacher quality, and how questions might be improved. Again, questions were revised based on this feedback.

Content and item structure was again reviewed and altered based on suggestions from the teachers and experts. The recommendations included the following: (1) Using the word “pupils” consistently throughout the measure, not “student,” (2) Shifting items to the domains where the fit was best, (3) Merging some items that were similar in meaning, (4) Deleting the items that were repetitious, and (5) Rephrasing items that were ambiguous. Recommendations three and four were not implemented as redundant items or items that lack independence would be identified through Rasch analysis. Moreover, according to Rasch theory, such items do not degrade the quality of measurement, but are inefficient because they do not provide unique information about
the respondents (Crocker & Algina, 2007). Following this, the 80-item measure was revised, and was ready to be piloted.

**Pilot Study of TQM (Including Revisions Made)**

In the pilot study, the revised survey was administered to teachers in the Kampala and Wakiso school districts in Uganda. These districts were selected because they consist of a variety of schools such as rural and urban, varying class sizes, and children of diverse backgrounds. One teacher in each of the schools in each district was asked to contact the other teachers and ask them to complete the TQM. Participants had approximately one month to respond (i.e., in October of 2011), and the survey took approximately 20 minutes to complete. Twenty teachers responded, with 11 (55%) males, and all professionally trained with either a Bachelor’s degree or higher. All teachers taught English, and were in urban \( n = 8; 40\% \) or rural \( n = 12; 60\% \) schools.

**Pilot Study Analysis**

Rasch Analysis was used to examine the psychometric properties of the 80-item TQM. Based on the diagnostics discussed, items were eliminated to produce the best possible model. The measure containing the original 80 items, after reviewer feedback, was run in WINSTEPS to examine Likert category functioning along with eliminating misfitting items. First, the rating scale diagnostics indicated that the first category was used infrequently (i.e., Observed Count = 16; 1%), and it was decided to combine the adjacent categories 0 and 1. Collapsing the 0 and 1 categories improved category definition in that the average measures and step calibrations became monotonic, and the probability curves showed that each category represented a distinct portion of the
underlying variable (Bond & Fox, 2007). Second, misfitting items were eliminated based on low point-measure correlations (i.e., < .15), high infit and outfit values (i.e., > 2.0), and examination of the variable map.

The misfit indices, point-measure correlations, and the variable map located several questionable items that appeared poorly fitting or redundant (i.e., 1, 16, 26, 28, 31, 37, 59, 64, 65, 66, 68, 74, and 77). Additionally, Pearson correlations were examined to assess the relationship between the items. Items that were found to be highly correlated (i.e., \( r > .70 \)) were removed (i.e., 23, 25, 24, 27, 29, 38, 39, 44, 46, and 60). Finally, a Reliability Analysis was run, and three items were removed, which would increase Coefficient Alpha (i.e., 3, 57, and 79). Therefore, 26 items were removed in total, leaving 54 items that were again examined using Rasch Analysis.

The 54-question TQM produced improved psychometric properties (see Table 3). For persons \((N = 20)\), separation was 4.05, and was 4.36 when the data have no misfit to the model. Item separation was 2.71. The person separation reliability estimate for these data was .94, and item reliability was .88. The model reliability had a value of .89 indicating that 89% of the variability among items is due to real item variance, and Coefficient Alpha was .96. With the exception of one item (i.e., 43), point measure correlations for this model were above .15. Infit and outfit measures less than 2.0 and were considered acceptable, and nine were at or above 2.0 (i.e., unstandardized or standardized). A few individual items appeared to have questionable misfit diagnostics, although not all of the indices (e.g., infit, outfit, point-measure correlations) suggested their elimination (see Table 4).
Table 3

Summary Statistics for the 54-Item Teacher Quality Measure (TQM) for Persons and Items

<table>
<thead>
<tr>
<th>SUMMARY OF 20 MEASURED PERSONS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RAW</strong></td>
<td><strong>COUNT</strong></td>
</tr>
<tr>
<td>MEAN</td>
<td>159.3</td>
</tr>
<tr>
<td>S.D.</td>
<td>27.2</td>
</tr>
<tr>
<td>MAX.</td>
<td>196.0</td>
</tr>
<tr>
<td>MIN.</td>
<td>93.0</td>
</tr>
<tr>
<td>REAL RMSE</td>
<td>.21</td>
</tr>
<tr>
<td>MODEL RMSE</td>
<td>.20</td>
</tr>
<tr>
<td>S.E. OF PERSON MEAN = .21</td>
<td></td>
</tr>
</tbody>
</table>

VALID RESPONSES: 99.6%
PERSON RAW SCORE-TO-MEASURE CORRELATION = .98 (approximate due to missing data)
CRONBACH ALPHA (KR-20) PERSON RAW SCORE RELIABILITY = .96 (approximate due to missing data)

<table>
<thead>
<tr>
<th>SUMMARY OF 54 MEASURED ITEMS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RAW</strong></td>
<td><strong>COUNT</strong></td>
</tr>
<tr>
<td>MEAN</td>
<td>59.0</td>
</tr>
<tr>
<td>S.D.</td>
<td>11.1</td>
</tr>
<tr>
<td>MAX.</td>
<td>75.0</td>
</tr>
<tr>
<td>MIN.</td>
<td>27.0</td>
</tr>
<tr>
<td>REAL RMSE</td>
<td>.36</td>
</tr>
<tr>
<td>MODEL RMSE</td>
<td>.34</td>
</tr>
<tr>
<td>S.E. OF ITEM MEAN = .14</td>
<td></td>
</tr>
</tbody>
</table>

UNEAN=.000 USCATE=1.000
ITEM RAW SCORE-TO-MEASURE CORRELATION = -.97 (approximate due to missing data)
1076 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 1921.29 with 1001 d.f. p=.0000
<table>
<thead>
<tr>
<th>ENTRY NUMBER</th>
<th>TOTAL COUNT</th>
<th>MEASURE S.E.</th>
<th>MODEL MNSQ</th>
<th>INFIT ZSTD</th>
<th>OUTFIT MNSQ</th>
<th>ZSTD</th>
<th>PTMEA</th>
<th>CORR.</th>
<th>EXACT OBS</th>
<th>MATCH EXPX</th>
<th>ITEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>34</td>
<td>20</td>
<td>1.94</td>
<td>24.290</td>
<td>4.31.426</td>
<td>4.41A0.00</td>
<td>25.0</td>
<td>38.3</td>
<td>043</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>31</td>
<td>20</td>
<td>2.10</td>
<td>24.189</td>
<td>2.31.396</td>
<td>5.68 .36</td>
<td>20.0</td>
<td>38.7</td>
<td>048</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>66</td>
<td>20</td>
<td>-0.52</td>
<td>35.224</td>
<td>2.61.58</td>
<td>1.6C .81</td>
<td>60.0</td>
<td>57.8</td>
<td>061</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>54</td>
<td>20</td>
<td>0.64</td>
<td>28.198</td>
<td>2.31.903</td>
<td>2.2D .76</td>
<td>20.0</td>
<td>36.8</td>
<td>030</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>27</td>
<td>20</td>
<td>2.33</td>
<td>24.187</td>
<td>2.12.161</td>
<td>1.4E .44</td>
<td>25.0</td>
<td>46.7</td>
<td>012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>35</td>
<td>20</td>
<td>1.88</td>
<td>24.187</td>
<td>2.41.644</td>
<td>1.4F .53</td>
<td>30.0</td>
<td>38.2</td>
<td>042</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>46</td>
<td>20</td>
<td>1.22</td>
<td>26.183</td>
<td>2.01.573</td>
<td>1.5G .65</td>
<td>35.0</td>
<td>45.2</td>
<td>035</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>55</td>
<td>18</td>
<td>-0.06</td>
<td>33.125</td>
<td>1.71.502</td>
<td>1.4H .36</td>
<td>50.0</td>
<td>51.5</td>
<td>058</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>61</td>
<td>20</td>
<td>0.02</td>
<td>31.156</td>
<td>1.61.374</td>
<td>1.7I .81</td>
<td>60.0</td>
<td>56.1</td>
<td>033</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>59</td>
<td>20</td>
<td>0.01</td>
<td>30.126</td>
<td>1.81.37</td>
<td>1.1K .58</td>
<td>50.0</td>
<td>50.6</td>
<td>039</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>59</td>
<td>20</td>
<td>-0.21</td>
<td>30.134</td>
<td>1.01.37</td>
<td>1.1K .58</td>
<td>50.0</td>
<td>50.6</td>
<td>039</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>44</td>
<td>20</td>
<td>1.35</td>
<td>25.123</td>
<td>0.81.304</td>
<td>0.9L .37</td>
<td>40.0</td>
<td>54.0</td>
<td>040</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>49</td>
<td>20</td>
<td>1.01</td>
<td>27.120</td>
<td>0.61.102</td>
<td>4N .77</td>
<td>35.0</td>
<td>46.6</td>
<td>017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>62</td>
<td>20</td>
<td>-0.08</td>
<td>32.125</td>
<td>1.61.14</td>
<td>1.1M .65</td>
<td>60.0</td>
<td>56.1</td>
<td>033</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>58</td>
<td>20</td>
<td>0.30</td>
<td>30.106</td>
<td>3.11.12</td>
<td>5O .26</td>
<td>40.0</td>
<td>48.2</td>
<td>032</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>70</td>
<td>20</td>
<td>-1.08</td>
<td>40.110</td>
<td>4.11.02</td>
<td>2P .59</td>
<td>50.0</td>
<td>63.6</td>
<td>062</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>68</td>
<td>20</td>
<td>-0.79</td>
<td>37.106</td>
<td>3.11.03</td>
<td>2Q .50</td>
<td>30.0</td>
<td>59.7</td>
<td>015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>74</td>
<td>20</td>
<td>-1.83</td>
<td>46.105</td>
<td>3.11.78</td>
<td>5R .67</td>
<td>80.0</td>
<td>75.3</td>
<td>022</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>60</td>
<td>20</td>
<td>-0.12</td>
<td>31.189</td>
<td>3.11.04</td>
<td>5S .46</td>
<td>50.0</td>
<td>56.0</td>
<td>050</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>49</td>
<td>20</td>
<td>1.01</td>
<td>27.100</td>
<td>1.11.90</td>
<td>2T .86</td>
<td>55.0</td>
<td>46.6</td>
<td>049</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>58</td>
<td>20</td>
<td>0.30</td>
<td>30.109</td>
<td>3.11.00</td>
<td>1U .17</td>
<td>50.0</td>
<td>48.2</td>
<td>032</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>69</td>
<td>20</td>
<td>-0.93</td>
<td>36.67</td>
<td>-0.995</td>
<td>0V .34</td>
<td>65.0</td>
<td>60.7</td>
<td>072</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>72</td>
<td>20</td>
<td>-1.42</td>
<td>45.93</td>
<td>-0.92</td>
<td>4W .62</td>
<td>60.0</td>
<td>65.8</td>
<td>080</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>44</td>
<td>19</td>
<td>1.16</td>
<td>27.82</td>
<td>-0.41</td>
<td>4X .73</td>
<td>47.4</td>
<td>54.1</td>
<td>047</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>71</td>
<td>20</td>
<td>-1.25</td>
<td>41.78</td>
<td>-0.91</td>
<td>1Y .35</td>
<td>55.0</td>
<td>64.8</td>
<td>067</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>54</td>
<td>20</td>
<td>0.64</td>
<td>28.99</td>
<td>-0.79</td>
<td>5Z .60</td>
<td>55.0</td>
<td>45.8</td>
<td>052</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 4**

**Item Statistics: Misfit Order (54 Measured Items)**
Examining the variable map (see Figure 2), in the 54-item TQM, most of the persons appeared above the mean on the vertical ruler. The items covered a range of -2.5 to 2.5 logits in difficulty, which was slightly broader than the range of about -1.5 to 3.0 for persons. This indicates that more “difficult” items may need to be added in the future. Also, at several points on the scale there were items at the same position on the vertical ruler, which indicates that these items may be redundant. The concerns presented above regarding model and item fit in this version of the TQM warranted further content analysis.

**Content Analysis of the 54-Item TQM**

As mentioned above, at several points on the scale, there were items at the same position on the vertical ruler, which indicates that these items were redundant. Therefore, a more detailed content analysis of some of the items was warranted. This was also needed as a 54-question survey was deemed too cumbersome to be completed by teachers. Some items appearing at the same logit value on the vertical ruler included Items 22, 32, and 78. Item 22 (i.e., “I treat each of my pupils fairly.”), Item 32 (i.e., “I build activities into the school day that support pupils working together.”), and Item 78 (i.e., “I demonstrate a respectful demeanour during the school day (e.g., classes, meetings).”), were removed as they were found to be closely related to Item 21’s content (i.e., “I have a respectful working relationship with my pupils”). That is, independent reviewers interpreted treating others “equally” as also having “respect” for them. Item 21 was retained because it encompassed the content of Items 22, 32, and 78. This item was
The distribution of person positions is on the left side of the vertical line and items on the right.

*Figure 2.* The map of persons and items for the 54-item Teacher Quality Measure (TQM).
also categorized in the Teacher Classroom Environment domain by both independent reviewers.

Independent reviewers found that Item 41 (i.e., “I allow pupils to practice after each step in the process of learning new material.”) and Item 42 (i.e., “I provide active practice of lesson material for all pupils.”) could be merged for a clearer meaning (i.e., “I present new lesson materials in small steps to allow pupils to practice after each step in the process of learning.”). Item 43 (i.e., “I give short quizzes at the beginning of the lesson.”) seemed vague. That is, the reason for administering quizzes needed to be well articulated in this item. Thus, it was rephrased to read: “I use short quizzes at the beginning of the lesson to check learning process.” Item 51 (i.e., “I provide feedback to pupils in a timely manner.”) was removed as it was a repetition of Item 52 (i.e., “I provide an immediate check for understanding by asking pupils to explain the facts that I have presented.”). Item 55 (i.e., “I use formal assessment strategies to evaluate the development of learners...”) as well as Item 56 (i.e., “I use informal assessment strategies to evaluate the development of learners...”) were also removed. The items did not conform to the construction or design of the other items, which may have contributed to some measurement error (e.g., they were too long).

Item 58 (i.e., “I display understanding of subject content matter in my area of specialization (i.e., in-depth knowledge of English.”), Item 61 (i.e., “I am available to pupils for consultation according to the school policies.”), Item 62 (i.e., “I am available to parents for consultation according to the school policies (e.g., during class meetings, visitations days and other days planned by the school).”), Item 71 (i.e., “I demonstrate
timeliness for assigned responsibilities from administrators as well items.”), and Item 75 (i.e., “I assume responsibility for the total school program.”) were removed as they seemed not to measure teacher quality directly. That is, the items seemed to focus more on obeying authority that may not translate to quality teaching, even though they seemed to be related to the domain (i.e., Teacher Professionalism).

Finally, Item 63 (i.e., “I maintain accurate records of pupils work/progress as well as non-instructional records (e.g., attendance).”) was removed as it was found to be vague in its construction. That is, to merely “maintain” accurate records of pupils without making good use of them, may not translate to quality teaching. Additionally, Item 69 (i.e., “I demonstrate ethical behavior in an educational setting (e.g., teaching, professional responsibilities.)”) was removed in favor of Item 68 (i.e., “I adhere to school procedures/rules.”), which seemed to include the content of Item 69. Overall, from the 54-item TQM, 16 more items were removed based on the above content analysis. This produced the 38-item TQM with acceptable psychometric properties, and is also a more portable and efficient measure compared to the earlier versions.

**Final TQM Analysis**

**Final TQM descriptive statistics.** The final TQM contained 38 items and was completed by 36 upper primary (i.e., 5th - 7th grade) teachers from the Wakiso School District in Uganda. The Likert scale for the measure ranges from 0 (i.e., Few) to 3 (i.e., Almost All), which corresponds to 0% to 25% of the time and 76% to 100% of the time, respectively. The total possible points on this measure is 114. The mean of the 38 items from the final analysis sample was 83.3 ($SD = 13.4$). The range was 70 with a minimum
score of 32 and a maximum score of 102. Histograms and skewness and kurtosis statistics revealed normally distributed data.

**Rasch summary statistics.** Prior to interpretation of the individual item and person data, appraisal of whether the data fit the model is required. Table 5 below presents overall information about whether the data showed acceptable fit to the model. The first overall statistic to consider is separation, the index of spread of the person positions or item positions. For persons, separation is 2.61 for the data at hand (real), and is 2.78 when the data have no misfit to the model (model). If separation is 1.0 or below, the items may not have sufficient breadth in position. In that case, it is critical to reconsider what having less and more of the construct means, and upon revision, add items that cover a broader range.

Item separation for these data was 2.28, a slightly smaller continuum than for persons. It is typical to find larger separation values for items than for persons. This is usually a function of the data having a smaller number of items and a larger number of people. However, in this analysis, there were 38 items and 36 people. Overall, the model had a separation value greater than 2, which indicates that true variability among items was larger than the amount of error variability. A larger item separation is preferable and perhaps a larger sample may remedy this.

Note that the mean for items was 0.0. The mean of the item logit position is always arbitrarily set at 0.0, similar to a standardized (z) score. The person mean here was 1.30, which suggests these items were easy, on average, for persons to “agree with.”
Table 5

Summary Statistics for the Final Teacher Quality Measure (TQM) for Persons and Items

The persons had a higher level of the trait than the items did. If the person mean was -1, -2, or +1 or +2, this would indicate that the items were potentially too hard or too easy for the sample. Thus, a person mean of 1.30 suggests that the items were too easy.

Following the examination of item means, mean infit and outfit for person and item mean squares were investigated. Mean infit and outfit is expected to be 1.0. For both, they were 1.04 and 1.01, respectively. Related to this, the mean standardized infit and outfit are expected to be 0.0. In the current model, they were .1 and .1 for persons,
and .1 and 0 for items. Overall, the data fit the model somewhat better than expected, which may signal some redundancy (i.e., possibly redundant items).

According to Bode and Wright (1999), the standard deviation of the standardized infit is an index of overall misfit for persons and items. Using 2.0 as a cut-off criterion, both persons (i.e., standardized infit SD = 1.0) and items (i.e., standardized infit SD = 1.3) showed little overall misfit. Here the data evidenced acceptable fit overall. This is in contrast to the overall Chi-Square test for this model, which was significant. This indicates that the Rasch model does not fit these items well \( \chi^2 = 2479.85, df = 1292, p = .000 \).

**Rasch response scale.** Table 6 below contains information about how the response scale was used. For these data, the response scale was 0 (Few), 1 (Many), 2 (Most), and 3 (Almost All). The step logit position is where a step indicates the transition from one rating scale category to the next (e.g., from a 2 to a 3). The “Observed Count” is the total of times the category was chosen from all items and all persons, whereas the "Observed Average" is the average of logit positions modeled in the category. It should increase by category value, and the current model demonstrated this. For example, persons responding with a 0 had an average measure (-.01) much lower than those responding with a 1 (.47).

There was no significant misfit for the categories as the misfit indices (i.e., mean square misfit) for the categories were below 1.5. "Sample Expected" is the best possible value of the average logit position for these data. These values should not be very different from the observed averages, and for these data they were not. The cut-off value
for infit and outfit mean squares is 1.0, and they were close to this value. Step (i.e., Structure) calibration is the logit calibrated difficulty of the step. This is shown in Figure 3 below. These values should increase with category value, and for the current model they did. It can also be observed that the standard error, which is a measure of uncertainty around the step calibration, were all near 0 for each step (i.e., .13, .08, and .06).

Table 6

*Summary of Category Structure (38 Measured Items) from the Teacher Quality Measure (TQM)*

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>OBSERVED</th>
<th>OBSVD</th>
<th>SAMPLE</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>STRUCTURE</th>
<th>CATEGORY</th>
<th>LABEL</th>
<th>SCORE</th>
<th>COUNT</th>
<th>AVGRE</th>
<th>EXPECT</th>
<th>MNSQ</th>
<th>MNSQ</th>
<th>CALIBRTN</th>
<th>MEASURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>74</td>
<td>5</td>
<td>.01</td>
<td>.13</td>
<td>1.14</td>
<td>1.35</td>
<td>NONE</td>
<td>-2.02</td>
<td>0</td>
<td></td>
<td>.01</td>
<td>.13</td>
<td>.14</td>
<td>.35</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>160</td>
<td>12</td>
<td>.47</td>
<td>.48</td>
<td>.99</td>
<td>.86</td>
<td>-59</td>
<td>-58</td>
<td>1</td>
<td></td>
<td>.47</td>
<td>.48</td>
<td>.99</td>
<td>.86</td>
<td>-59</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>392</td>
<td>29</td>
<td>.98</td>
<td>1.03</td>
<td>.90</td>
<td>.86</td>
<td>-14</td>
<td>.54</td>
<td>2</td>
<td></td>
<td>.98</td>
<td>1.03</td>
<td>.90</td>
<td>.86</td>
<td>-14</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>741</td>
<td>54</td>
<td>1.79</td>
<td>1.77</td>
<td>1.02</td>
<td>1.02</td>
<td>.73</td>
<td>2.08</td>
<td>3</td>
<td></td>
<td>1.79</td>
<td>1.77</td>
<td>1.02</td>
<td>1.02</td>
<td>.73</td>
</tr>
<tr>
<td>MISSING</td>
<td>1</td>
<td>0</td>
<td>.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.57</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>STRUCTURE</th>
<th>SCORE-TO-MEASURE</th>
<th>50% CUM.</th>
<th>COHERENCE</th>
<th>PROBABILITY</th>
<th>M-&gt;C</th>
<th>C-&gt;M</th>
<th>RMSR</th>
<th>ESTIM.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>NONE</td>
<td>( -2.02 ) INF</td>
<td>-1.35</td>
<td>63%</td>
<td>7%</td>
<td>1.6589</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-59</td>
<td>.13</td>
<td>-1.53</td>
<td>-1.00</td>
<td>52%</td>
<td>2%</td>
<td>.03</td>
<td>1.0273</td>
<td>.83</td>
</tr>
<tr>
<td>2</td>
<td>-14</td>
<td>.08</td>
<td>.54</td>
<td>-1.34</td>
<td>40%</td>
<td>64%</td>
<td>.4754</td>
<td>1.04</td>
<td>.95</td>
</tr>
<tr>
<td>3</td>
<td>.73</td>
<td>.06</td>
<td>.13</td>
<td>INF</td>
<td>78%</td>
<td>66%</td>
<td>.5755</td>
<td>1.07</td>
<td>3</td>
</tr>
</tbody>
</table>

M->C - Does Measure Imply Category? C->M = Does Category Imply Measure?

**Rasch probability curves.** A final way of examining step use is via probability curves. These curves display the likelihood of category selection (Y-axis) by the person-minus-item measure (X-axis). If all categories are utilized as expected, each category value will be the most likely at some point on the continuum (i.e., as shown below in Figure X), and there will be no category inversions where a higher category is more
likely at a lower point than a lower category. For these data, all the categories were utilized; however, it can be seen that Category 1 is almost eclipsed by 0 and 2, which might indicate that respondents are not using the response scale as intended, or that the “Many” category might not be necessary.

**Figure 3.** Category probabilities (i.e., probability curves) indicating the probability of a response for the 38-item Teacher Quality Measure (TQM). These curves display the likelihood of category selection (Y-axis) by the person-minus-item measure (X-axis).

**Rasch item misfit diagnostics.** Table 7 below presents item misfit diagnostics. Measure is the logit position of the item, with error being the standard error of measurement for the item. Item fit for the model was determined by the infit and outfit diagnostics and the point measure correlations. Infit “…is a t standardized information-weighted mean square statistic, which is more sensitive to unexpected behavior affecting
responses to items near the person's measure level” (Linacre, 2009, p. 252). Outfit “… is a standardized outlier-sensitive mean square fit statistic, more sensitive to unexpected behavior by persons on items far from the person's measure level” (Linacre, 2009, p.252). A point measure correlation is the correlation between the item score and the measure, which should be positive (Linacre, 2009).

Infit and outfit values of less than 2 were considered acceptable (Smith, 2004). Infit for all items on this measure were either less than 2 or extremely close to 2 (e.g., Item 26 [2.02] and Item 38 [2.04]), and were therefore acceptable. Most outfit measures were less than 2 as well. Item 15 (i.e., “I treat each of my pupils equally.”) had an outfit of 2.49, and Item 26 (i.e., “I do not teach to a variety of learning styles.”) had an outfit of 2.58. A vast majority of the sample endorsed the highest category on the scale for these items. Thus, these items were very easy to endorse, which could account for the high outfit value.

A correlation below .15 indicates a potentially misfitting item, and the values are preferably between .3 and .5. Point measure correlations for this model ranged from .00 and .60. Unsurprisingly, the two lowest correlations were with the items listed above. As mentioned above, a point measure correlation below .15 indicates a potentially misfitting item, and these two items were already suspicious due to their large outfit values. None of the other items had correlations below .15.
### Table 7

**Item Statistics: Misfit Order (38 Measured Items)**

<table>
<thead>
<tr>
<th>ENTRY NUMBER</th>
<th>TOTAL SCORE</th>
<th>TOTAL COUNT</th>
<th>MEASURE</th>
<th>MODEL S.E.</th>
<th>INFIT MNSQ</th>
<th>ZSTD</th>
<th>OUTFIT MNSQ</th>
<th>ZSTD</th>
<th>PT-MEASURE</th>
<th>CORR. EXP.</th>
<th>EXACT MATCH</th>
<th>OBSN</th>
<th>EXP%</th>
<th>ITEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>76</td>
<td>36</td>
<td>.45</td>
<td>.21 .02 .2</td>
<td>3.6 .25 .4</td>
<td>4.2</td>
<td>A .00 .45</td>
<td>27.8</td>
<td>44.5 Q26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>94</td>
<td>36</td>
<td>-.54</td>
<td>.28 .16 .2</td>
<td>1.8 .24 .1</td>
<td>2.8</td>
<td>B .03 .32</td>
<td>52.8</td>
<td>66.5 Q15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>95</td>
<td>36</td>
<td>-.62</td>
<td>.29 .04 .2</td>
<td>2.5 .17 .6</td>
<td>1.6</td>
<td>C .21 .31</td>
<td>61.1</td>
<td>67.6 Q38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>89</td>
<td>36</td>
<td>-.19</td>
<td>.25 .19 .2</td>
<td>2.7 .17 .4</td>
<td>1.6</td>
<td>D .32 .37</td>
<td>38.9</td>
<td>53.7 Q18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>71</td>
<td>36</td>
<td>2.32</td>
<td>.24 .69 .5</td>
<td>1.6 .66 .2</td>
<td>2.2</td>
<td>E .32 .57</td>
<td>27.8</td>
<td>43.4 Q25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>71</td>
<td>36</td>
<td>.66</td>
<td>.20 .12 .1</td>
<td>1.0 .12 .2</td>
<td>1.7</td>
<td>F .36 .47</td>
<td>38.9</td>
<td>42.9 Q17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>36</td>
<td>-1.14</td>
<td>.36 .43 .1</td>
<td>1.1 .97 .9</td>
<td>1.1</td>
<td>G .32 .25</td>
<td>83.3</td>
<td>80.3 Q4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>86</td>
<td>36</td>
<td>-.02</td>
<td>.23 .31 .1</td>
<td>1.2 .23 .8</td>
<td>.8</td>
<td>H .32 .39</td>
<td>55.6</td>
<td>51.7 Q10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>98</td>
<td>36</td>
<td>-.90</td>
<td>.33 .15 .5</td>
<td>.8 .89 .8</td>
<td>-1</td>
<td>I .39 .28</td>
<td>77.8</td>
<td>75.8 Q2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>84</td>
<td>36</td>
<td>.09</td>
<td>.23 .12 .2</td>
<td>.9 .14 .6</td>
<td>1.1</td>
<td>J .36 .40</td>
<td>38.9</td>
<td>50.6 Q8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>79</td>
<td>36</td>
<td>.32</td>
<td>.21 .10 .4</td>
<td>.3 .15 .6</td>
<td>.6</td>
<td>K .36 .43</td>
<td>38.9</td>
<td>46.4 Q27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>92</td>
<td>36</td>
<td>-.39</td>
<td>.27 .10 .0</td>
<td>.1 .10 .4</td>
<td>.4</td>
<td>L .32 .34</td>
<td>61.1</td>
<td>63.0 Q6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>97</td>
<td>36</td>
<td>-.80</td>
<td>.31 .09 .4</td>
<td>.4 .93 .0</td>
<td>.0</td>
<td>M .30 .29</td>
<td>72.2</td>
<td>72.0 Q35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>63</td>
<td>36</td>
<td>.96</td>
<td>.19 .10 .6</td>
<td>.4 .10 .4</td>
<td>.4</td>
<td>N .42 .50</td>
<td>44.4</td>
<td>41.1 Q8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>80</td>
<td>36</td>
<td>.28</td>
<td>.21 .10 .8</td>
<td>.4 .10 .1</td>
<td>.1</td>
<td>O .52 .43</td>
<td>50.0</td>
<td>46.4 Q13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>90</td>
<td>36</td>
<td>-.25</td>
<td>.25 .99 .1</td>
<td>.9 .93 .1</td>
<td>-1</td>
<td>P .26 .36</td>
<td>55.6</td>
<td>57.2 Q36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>75</td>
<td>36</td>
<td>.50</td>
<td>.20 .97 .0</td>
<td>.0 .89 .3</td>
<td>-1</td>
<td>Q .52 .45</td>
<td>41.7</td>
<td>44.3 Q24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>82</td>
<td>36</td>
<td>.19</td>
<td>.22 .97 .0</td>
<td>.0 .85 .4</td>
<td>-4</td>
<td>R .49 .41</td>
<td>58.3</td>
<td>48.1 Q3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>69</td>
<td>36</td>
<td>.73</td>
<td>.19 .95 .2</td>
<td>.8 .85 .6</td>
<td>-6</td>
<td>S .55 .48</td>
<td>52.8</td>
<td>42.9 Q21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>70</td>
<td>31</td>
<td>.62</td>
<td>.20 .89 .3</td>
<td>.5 .90 .3</td>
<td>-3</td>
<td>S .44 .47</td>
<td>60.0</td>
<td>43.1 Q7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>91</td>
<td>36</td>
<td>-.32</td>
<td>.26 .88 .3</td>
<td>.7 .75 .6</td>
<td>-6</td>
<td>T .44 .35</td>
<td>55.6</td>
<td>59.2 Q20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>63</td>
<td>36</td>
<td>.96</td>
<td>.19 .86 .6</td>
<td>.6 .84 .7</td>
<td>-4</td>
<td>U .56 .50</td>
<td>44.4</td>
<td>41.1 Q19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>95</td>
<td>36</td>
<td>-.62</td>
<td>.28 .86 .3</td>
<td>.7 .79 .4</td>
<td>-4</td>
<td>V .35 .31</td>
<td>61.1</td>
<td>67.6 Q33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>102</td>
<td>36</td>
<td>-.14</td>
<td>.42 .82 .3</td>
<td>.3 .84 .1</td>
<td>-1</td>
<td>W .24 .22</td>
<td>60.6</td>
<td>44.4 Q29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>78</td>
<td>36</td>
<td>.97</td>
<td>.21 .75 .1</td>
<td>.1 .84 .5</td>
<td>.1</td>
<td>X .34 .44</td>
<td>52.8</td>
<td>46.2 Q1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>96</td>
<td>36</td>
<td>.71</td>
<td>.30 .84 .4</td>
<td>.7 .70 .6</td>
<td>-6</td>
<td>Y .33 .30</td>
<td>66.7</td>
<td>68.8 Q28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>90</td>
<td>36</td>
<td>-.25</td>
<td>.25 .82 .6</td>
<td>.6 .67 .9</td>
<td>-1</td>
<td>Z .43 .36</td>
<td>66.7</td>
<td>57.2 Q37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>69</td>
<td>36</td>
<td>.73</td>
<td>.19 .82 .8</td>
<td>.8 .80 .8</td>
<td>-8</td>
<td>A .51 .48</td>
<td>61.1</td>
<td>42.9 Q11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>96</td>
<td>36</td>
<td>-.71</td>
<td>.30 .82 .4</td>
<td>.7 .73 .5</td>
<td>.1</td>
<td>B .41 .30</td>
<td>75.0</td>
<td>66.8 Q34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>92</td>
<td>36</td>
<td>-.39</td>
<td>.27 .76 .7</td>
<td>.6 .64 .9</td>
<td>-3</td>
<td>C .52 .34</td>
<td>66.7</td>
<td>63.0 Q23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>92</td>
<td>36</td>
<td>-.39</td>
<td>.27 .75 .8</td>
<td>.6 .66 .9</td>
<td>-9</td>
<td>D .42 .34</td>
<td>61.1</td>
<td>63.0 Q14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>90</td>
<td>36</td>
<td>-.25</td>
<td>.25 .75 .9</td>
<td>.7 .75 .7</td>
<td>-7</td>
<td>E .47 .36</td>
<td>62.1</td>
<td>57.2 Q12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>85</td>
<td>36</td>
<td>-.03</td>
<td>.30 .82 .4</td>
<td>.7 .73 .5</td>
<td>.1</td>
<td>F .41 .30</td>
<td>75.0</td>
<td>66.8 Q9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>81</td>
<td>36</td>
<td>.23</td>
<td>.22 .68 .1</td>
<td>1.4 .64 .1</td>
<td>1.3</td>
<td>G .49 .42</td>
<td>55.6</td>
<td>48.0 Q22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>86</td>
<td>36</td>
<td>-.02</td>
<td>.23 .66 .4</td>
<td>1.4 .60 .1</td>
<td>1.3</td>
<td>H .57 .39</td>
<td>63.9</td>
<td>51.7 Q31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>73</td>
<td>36</td>
<td>.69</td>
<td>.27 .62 .1</td>
<td>1.4 .67 .1</td>
<td>1.2</td>
<td>I .49 .34</td>
<td>72.2</td>
<td>63.0 Q17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>92</td>
<td>36</td>
<td>.39</td>
<td>.27 .62 .1</td>
<td>1.4 .67 .1</td>
<td>1.2</td>
<td>J .49 .34</td>
<td>72.2</td>
<td>63.0 Q30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>74</td>
<td>36</td>
<td>.64</td>
<td>.23 .58 .2</td>
<td>2.2 .56 .3</td>
<td>1.9</td>
<td>A .56 .46</td>
<td>55.6</td>
<td>44.2 Q2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Mean:** 36.0  
**S.D.:** 13.2
Rasch map of persons and items. The map of persons and items is shown in Figure 4 below. The distribution of person positions is on the left side of the vertical line and items on the right. Each "X" represents one person in this figure. "M" marks the person and item mean; "S" is one standard deviation away from the mean; and "T" is two standard deviations away from the mean. To determine variability, item measure values were investigated using the item/person map for this model. The degree to which these items are targeted at the teachers was investigated. As seen below, the scale appeared to be applicable for its purposes. The items were normally distributed; however, some items (e.g., Items 25 and 29) were separated from the others on the far ends of the scale. A few items appear at the same logit value, indicating some redundancy. Fortunately, there were no large gaps in between the items, which could indicate that more items need to be added to address the full range (i.e., content domain) of the construct.

The map shows many persons appearing above where the items are targeted. The items covered a range from -1.5 to 2 logits in difficulty, which is narrower than the range of about -.5 to 4 for persons. That is, most teachers found most items easy to endorse (i.e., endorsing the highest category on the Likert scale), with the exception of Item 25 (i.e., “I have pupils prepare a written summary of the day’s lesson”). Overall, the above evidence indicates that more difficult items to endorse may need to be added in future studies to fill in the range of the construct measured, or some items should be considered for elimination based on misfit diagnostics or placement on the variable map.
The distribution of person positions is on the left side of the vertical line and items on the right.

*Figure 4.* The map of persons and items for the 38-item Teacher Quality Measure (TQM).
Research Question 2A

Research Question 2A (i.e., “What is the relationship between a pupil’s sex and achievement on the PLE (i.e., 7th grade) in Uganda?”) was analyzed using a Two-Factor Analysis of Covariance (ANCOVA). This was conducted to determine the relationship between a pupil’s sex and achievement on the PLE in Uganda, while controlling for pupils’ prior mathematics and English achievement/knowledge. To examine this research question, data were used from multiple sources. Student data (i.e., sex, SES, PLE pretest scores in mathematics and English, final mathematics and English PLE scores) were gathered from the Wakiso district.

Mathematics Descriptives

Before conducting the analysis, outliers were removed ($n = 49$). That is, all PLE pretest and posttest scores with $z$ scores $\geq 1.96$ for were removed. Thus, the final analysis sample contained 435 boys and 468 girls, and 208 boarding school students and 695 day school students ($N = 903$). For the pretest mathematics in these data, the girls had higher scores ($M = 7.42, SD = 2.02$) compared to the boys ($M = 6.66, SD = 2.08$). Additionally, for the posttest, the boys’ scores were lower ($M = 4.28, SD = 2.31$) than the girls ($M = 5.44, SD = 2.17$). Furthermore, boarding schools attained higher scores on average ($M = 5.54, SD = 2.04$) for the pretest compared to the posttest ($M = 2.62, SD = 1.42$). Finally, day schools also had higher means at the pretest ($M = 7.50, SD = 1.88$) compared to the posttest ($M = 5.56, SD = 2.09$).
Mathematics Assumptions

The assumption of Independence was tested via a scatterplot of residuals. A random display of points of the residuals against values of independent variables (i.e., sex and SES) provided evidence that the assumption of independent errors was met. Normality was tested on the covariate (i.e., pretest mathematics test scores), the dependent variable (i.e., posttest mathematics test scores), and separately in each group of the independent variables (male and female and day and boarding). For all the above, the assumption was tested via examinations of Shapiro-Wilk (S-W), skewness and kurtosis, histograms, and Q-Q plots.

Z scores for skewness (S) and kurtosis (K) were calculated by taking the statistic divided by the standard error and comparing the value to ± 1.96. Overall pretest mathematics (S = -7.84, K = -6.25) and posttest mathematics (S = 1.94, K = -6.54), as well as boys pretest (S = -2.80, K = -5.29) and girls pretest (S = -8.69, K = -2.12) indicated significant skewness and kurtosis in most instances. The same was found for boys posttest (S = 4.09, K = 3.88), but not girls posttest mathematics (S = -5.8, K = 4.06). For boarding schools and day schools, the boarding pretest (S = 2.88, K = -2.90) and day school pretest (S = -11.27, K = -.83), and boarding posttest (S = 8.23, K = 5.74) and day schools posttest (S = -.31, K = -5.14) were significantly skewed and kurtotic in most groups. The S-W test for normality yielded significant results across all variables (p < .05), which indicates that normality has been violated. However, histograms and Q-Q plots suggested relatively normal distributional shapes across the variables and groups. In most cases, the pretest was negatively skewed with many higher scores (i.e., in this case,
higher is poorer performance), compared to the posttest. This scenario is expected as pupils tend to improve their scores from pretest to posttest. Overall, except for the S-W test, which tends to be conservative, there is evidence of normality.

Linearity of the dependent variable with the covariate was examined with scatterplots, both overall and by group. Overall, the scatterplots suggested a positive linear relationship. Homogeneity of Regression Slopes was examined by similar regression lines evidenced in the scatterplots of the dependent variable and covariate by the independent variable (reported earlier as evidence for linearity). This assumption was confirmed by a nonstatistically significant interaction \( (F (1, 898) = .410, p = .522) \). The Homogeneity of Variance assumption was not satisfied \((p < .05)\), which may be a product of the disparate group sizes for SES. Additionally, research suggests that violation of this assumption is minimal when the groups of the independent variable are approximately equal in size (Harwell, 2003), as was the case with sex.

**Mathematics ANCOVA**

A Two-Factor ANCOVA was conducted to assess the effect of Sex (and SES) on student mathematics achievement controlling for mathematics pretest. The IVs were Sex and SES. The DV was scores on the PLE mathematics exam (i.e., the posttest) that is administered following completion of 7th grade. Scores on a mock mathematics test administered at the beginning of the 7th grade were used as the covariate (i.e., the pretest). The results of the ANCOVA suggest a statistically significant main effect of the covariate on the DV \((F = 478.64, df= 1, 898, p< .001)\). There were also statistically significant main effects of Sex \((F = 36.96, df= 1, 898, p< .001)\) and SES \((F = 166.35, df= 1, 898, p<\)
.001) on the posttest adjusting for the pretest. There was also a nonsignificant interaction effect between Sex and SES (\(p = .124\)). The effect sizes for Sex and SES were small (partial \(\eta^2 = .08\)) and large (partial \(\eta^2 = .14\)), respectively. Observed power was strong for both main effects.

The results indicate that there are significant differences between boys and girls on mathematics achievement controlling for prior mathematics ability. Specifically, the adjusted means indicate that boys have higher mathematics achievement on average (\(M = 4.10, SE = .09\)) compared to girls (\(M = 4.61, SE = .12\)). Note that the lower mean indicates better performance, and a higher mean indicates poorer performance on the PLE exams. (i.e., 1 is the highest score and 9 is the lowest score).

The results were analyzed for SES, even though this was not considered the focus of this research question. The findings demonstrate that there are significant differences between day and boarding schools on mathematics achievement controlling for prior mathematics ability. Specifically, the results show that the mathematics achievement adjusted mean for boarding schools (i.e., indicating a higher SES; \(M = 3.38, SE = .14\)) is higher compared to the day schools (i.e., indicating a lower SES; \(M = 5.33, SE = .07\)). That is, the pupils with a higher SES have a higher level of math achievement than their counterparts who have a lower SES controlling for prior mathematics ability.

**English Descriptives**

Before conducting the analysis, outliers were removed (\(n = 100\)). That is, all PLE pretest and posttest scores with \(z\) scores \(\pm 1.96\) for were removed. Thus, the final analysis sample contained 413 boys and 439 girls, and 225 boarding students and 627 day
students ($N = 852$). Overall, the pretest scores were higher ($M = 6.30$, $SD = 2.20$) than the posttest scores ($M = 3.45$, $SD = 1.58$). For the English pretest in these data, girls scores were higher on average ($M = 6.78$, $SD = 2.05$) compared to the boys ($M = 5.78$, $SD = 2.24$). For the posttest, the girls’ scores were also higher ($M = 3.69$, $SD = 1.59$) compared to the boys ($M = 3.20$, $SD = 1.55$). Furthermore, boarding schools attained higher means ($M = 4.47$, $SD = 1.72$) for the pretest compared to the posttest ($M = 2.12$, $SD = .77$).

Finally, day schools attained higher means ($M = 6.95$, $SD = 1.97$) at the pretest than ($M = 3.93$, $SD = 1.53$) at the posttest.

**English Assumptions**

The assumption of Independence was tested via a scatterplot of residuals. A relatively random display of points of the residuals against values of independent variables (i.e., sex and SES) provided evidence of Independence. Normality was tested on the covariate (i.e., pretest English scores), the dependent variable (i.e., posttest English scores), and also separately in each group of the independent variables (male and female and day and boarding). For all the above, the assumption was tested via examinations of S-W, skewness and kurtosis, histograms, and Q-Q plots.

$Z$ scores for skewness as well as kurtosis were again calculated. Pretest English ($S = -.10$, $K = -5.07$) and posttest English ($S = .13$, $K = -6.35$), and boys pretest ($S = -1.11$, $K = -4.23$) and girls pretest ($S = -4.62$, $K = -2.17$) were found to be not significantly skewed, but slightly kurtotic. For boarding schools and day schools, the boarding pretest ($S = 1.41$, $K = -1.25$) and day school pretest ($S = -7.06$, $K = -.130$), and the boarding posttest ($S = 11.04$, $K = 58.06$) and day schools posttest ($S = 9.35$, $K = 3.75$) were mostly
all significantly skewed or kurtotic. Additionally, the S-W test for normality yielded significant results across all variables and groups ($p < .05$). However, the histograms and Q-Q plots suggested relatively normal distributional shapes across all the variables and groups. Again, this was expected from pretest to posttest. Overall, there is some evidence that normality has been met.

Linearity of the dependent variable with the covariate was examined with scatterplots, both overall and by group. Overall, the scatterplots suggested a positive linear relationship. Homogeneity of Regression Slopes was suggested by similar regression lines evidenced in the scatterplots of the dependent variable and covariate by the independent variable (reported earlier as evidence for linearity). This assumption was confirmed by a nonstatistically significant interaction ($F (1, 890) = .124, p = .725$). The homogeneity of variance was not satisfied ($p < .05$). However, research suggests that violation of this assumption is minimal when the groups of the independent variable are approximately equal in size (Harwell, 2003), as was the case with sex.

**English ANCOVA**

A Two-Factor ANCOVA was conducted to assess the effect of Sex (and SES) on student English achievement controlling for English pretest. Again, the IVs were Sex and SES. The DV was scores on the PLE English exam (i.e., the posttest) that is administered following completion of the 7th grade. Scores on a mock English test administered at the beginning of the 7th grade were used as the covariate (i.e., the pretest). The results of the ANCOVA suggest a statistically significant main effect of the covariate on the DV ($F = 384.45$, $df = 1, 847$, $p < .001$). There was also a statistically significant main effect of SES
(F = 96.595, df = 1, 847, p < .001), but not sex (F = 1.621, df = 1, 847, p = .203) on the posttest adjusting for the pretest. There was a nonsignificant interaction effect between Sex and SES (p = .136). The effect size for SES was medium to large (partial $\eta^2 = .102$), and observed power was strong.

The results indicated that there is not a significant difference between boys and girls on English achievement controlling for prior English ability. The adjusted means indicated that boys have slightly lower English achievement on average ($M = 3.23, SE = .08$) compared to girls ($M = 3.19, SE = .07$), but again, there was no statistical difference. Note that the lower mean indicates better performance, and a higher mean indicates poorer performance on the PLE exams. (i.e., 1 is the highest score and 9 is the lowest score).

The results were also reported for SES. The findings demonstrate that there is a significant relationship between SES and English achievement controlling for prior English ability. Specifically, the English achievement adjusted mean for boarding schools (i.e., indicating a higher SES; $M = 2.66, SE = .10$) is higher compared to the day schools (i.e., indicating a lower SES; $M = 3.76, SE = .05$). That is, the pupils from high SES schools have a higher level of English achievement than their counterparts from lower SES schools controlling for prior English ability.

Research Question 2B

Research Question 2B stated, “What is the relationship between teacher quality and a pupil’s achievement on the PLE (i.e., 7th grade) in Uganda?” This was analyzed using a Two-Level Hierarchical Linear Model (HLM). HLM was used to investigate the
relationship between teacher quality and pupils’ achievement on PLE achievement in mathematics and English controlling for prior achievement/knowledge. Student scores in English and mathematics on the PLE, as well as the teacher scores from the newly developed TQM were used.

The HLM modeling procedure had three steps. First, the analysis included the null model (i.e., the One-Way Random Effects ANOVA) with only the pupil-level outcome variable (i.e., mean PLE achievement). Second, a pupil-level model was developed (i.e., the Unconditional Model) including the effects of prior achievement on the DV (i.e., PLE scores). Teacher variables were added to the pupil model at the third step (i.e., the Contextual Model). The above is the common model-building procedure used in HLM, although the third step (i.e., the Contextual Model) is the one that specifically addresses Research Question 2B. As mentioned earlier, the teacher-specific intercepts, which are treated as random at the second level, were adjusted (Raudenbush & Bryk, 2002) with grand-mean centering. The method of estimation used was Restricted Maximum Likelihood (REML).

**Mathematics Descriptives**

An outlier analysis was conducted on all variables in the full model and none were found (i.e., $z > \pm 1.96$). Thus, the final analysis sample contained all 952 students. For the mathematics data, the descriptive statistics for the level 1 variables included higher pretest means of 6.78 ($SD = 2.35$) compared to the posttest of 4.49 ($SD = 2.59$). The level 2 variable (i.e., the TQM) sample contained 18 teachers with an average TQM score of 90.47 ($SD = 12.08$). The scores ranged from 50 to 110.
**Mathematics Assumptions**

Normality was tested for all variables (i.e., pretest mathematics, posttest mathematics, as well as the TQM). Histograms and skewness and kurtosis statistics suggested that normality was upheld.

**Mathematics One-Way Random Effects ANOVA.** Table 8 below shows the results of One-Way Random Effects ANOVA Model (i.e., the Empty Model). The average student mathematics PLE mean was statistically different from zero ($\gamma_{00} = 4.76$, $t = 10.98$, $df = 17$, $p = .000$). Considerable variation in the student mathematics means still exists ($\tau_{00} = 3.05$, $X^2 = 11.74.77$, $df = 17$, $p = .000$). The proportion of the total variance in mathematics PLE achievement that can be attributed to the teacher was 48%. That is,

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient (SE)</th>
<th>$t$ (df)</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model for Mean Teacher PLE Math Scores ($\beta_0$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept ($\gamma_{00}$)</td>
<td>4.76 (.44)</td>
<td>10.98 (721)</td>
<td>.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Variance</th>
<th>$df$</th>
<th>$X^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variation in Teacher Means ($\tau_{00}$)</td>
<td>3.46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variation within Teachers ($\sigma^2$)</td>
<td>3.05</td>
<td>17</td>
<td>1174.77</td>
<td>.000</td>
</tr>
</tbody>
</table>

*Note.* Deviance (REML) = 3832.86; 2 estimated parameters
48% of the variability in mathematics PLE achievement can be attributed to teacher quality. Based on the significant amount of unexplained variability, additional Level 1 predictors (i.e., pretest scores) were added to try and reduce the variation within students, as well as adding other Level 2 variables to explain between student differences in the following models.

**Mathematics Unconditional Model**

Table 9 below shows the results of the Unconditional Model with the mathematics pretest as the sole predictor at Level 1 and no Level 2 variables. After including the pretest (i.e., grand-mean centered) as a predictor of mathematics PLE achievement,

Table 9

*Unconditional Model with the Mathematics Primary Leaving Examination (PLE)*

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient (SE)</th>
<th>t (df)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model for Mean Teacher PLE Math Scores (β₀)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept (γ₀₀)</td>
<td>4.61 (.40)</td>
<td>11.66 (17)</td>
<td>.000</td>
</tr>
<tr>
<td>Model for Pretest Slopes (β₁)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept (γ₁₀)</td>
<td>.36 (.06)</td>
<td>8.31 (17)</td>
<td>.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Variance</th>
<th>df</th>
<th>X²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variation in Teacher Means (τ₀₀)</td>
<td>2.91</td>
<td>17</td>
<td>738.08</td>
<td>.000</td>
</tr>
<tr>
<td>Variation in Pretest Slopes (τ₁₁)</td>
<td>.03</td>
<td>17</td>
<td>57.13</td>
<td>.000</td>
</tr>
<tr>
<td>Variation within Teachers (σ²)</td>
<td>2.58</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Deviance (REML) = 3679.85; 4 estimated parameters
within-teacher variability in mathematics PLE scores was reduced by 47%, relative to the One-Way Random Effects ANOVA Model. Overall, mathematics PLE scores across teachers was still significantly different from zero ($\gamma_{00} = 4.61, t = 11.66, df = 17, p = .000$). On average, across teachers, student pretest scores were positively and statistically significantly related to mathematics PLE achievement within teachers ($\gamma_{10}$). The average effect (i.e., slopes) across teachers for the pretest is represented as an increase of .36 points on the mathematics PLE for every one-unit increase in pretest scores.

Statistically significant differences (variability) in the 18 teacher means still exists ($\tau_{00}$). There was significant variability in both the intercepts and slopes. Statistically significant variability in the mathematics PLE means still exists after considering the pretest ($\tau_{00} = 2.91, \chi^2 = 738.41, df = 17, p = .000$), as well as statistically significant variability in the effect of the pretest (i.e., the slopes) across teachers ($\tau_{11} = .03, \chi^2 = 57.13, df = 17, p = .000$). Based on the significant amount of unexplained variability, a Level 2 predictor (i.e., teacher TQM scores) was added to explain the variation between students.

**Mathematics Contextual (Full) Model**

After adding the Level 2 predictor (i.e., teacher TQM scores grand-mean centered) to the model, 41% of the variance in mathematics PLE achievement can be explained by teacher TQM scores (see Table 10 below). Additionally, 33% of the variance in the effect of the pretest (i.e., the slopes) can be explained by teacher TQM scores. Mathematics PLE scores across teachers was still significantly different from zero ($\gamma_{00} = 4.62, t = 15.55, df = 16, p = .000$). That is, the average teacher has mean
mathematics PLE scores of 4.10 for their students. Teacher TQM score had a statistically significant influence on the intercept for mathematics PLE achievement. The effect of teacher TQM scores on mean PLE mathematics achievement is negative and statistically significant ($\gamma_{01} = -.08$, $t = -3.74$, $df = 16$, $p = .002$). The value of -.08 can be interpreted as the decrease in a teacher’s mean mathematics PLE achievement to be expected for a one-unit increase in teacher TQM scores, on average. The average teacher, therefore, is predicted to have a mean mathematics PLE achievement, on average, of 4.62 for their students, and with every one-point increase in TQM scores (i.e., increasing teacher quality), the mean mathematics PLE scores would decrease by .08 points. Thus, more quality teachers have students with better mathematics PLE scores on average.

Table 10

Contextual (Full) Model with the Mathematics Primary Leaving Examination (PLE)

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient (SE)</th>
<th>t (df)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model for Mean Teacher PLE Math Scores ($\beta_0$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept ($\gamma_{00}$)</td>
<td>4.62 (.30)</td>
<td>15.55 (16)</td>
<td>.000</td>
</tr>
<tr>
<td>Teacher Quality Measure (TQM; $\gamma_{01}$)</td>
<td>-.08 (.02)</td>
<td>-3.74 (16)</td>
<td>.002</td>
</tr>
<tr>
<td>Model for Pretest Slopes ($\beta_1$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept ($\gamma_{10}$)</td>
<td>.36 (.04)</td>
<td>8.82 (16)</td>
<td>.000</td>
</tr>
<tr>
<td>Teacher Quality Measure (TQM; $\gamma_{11}$)</td>
<td>-.01 (.01)</td>
<td>-2.17 (16)</td>
<td>.045</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Variance</th>
<th>df</th>
<th>$X^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variation in Teacher Means ($\tau_{00}$)</td>
<td>1.72</td>
<td>16</td>
<td>336</td>
<td>.000</td>
</tr>
<tr>
<td>Variation in Pretest Slopes ($\tau_{11}$)</td>
<td>.02</td>
<td>16</td>
<td>40.85</td>
<td>.000</td>
</tr>
<tr>
<td>Variation within Teachers ($\sigma^2$)</td>
<td>5.58</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Deviance (REML) = 3687.50; 4 estimated parameters
Again, the effect of the pretest on mathematics PLE achievement is positive on average and statistically significant ($\gamma_{10} = .36, t = 8.82, df = 16, p = .000$). Thus, there was a statistically significant effect of the pretest slope (i.e., predicting the mathematics PLE scores) across students and teachers. Specific to the research question, the impact of the TQM on mathematics PLE scores controlling for the pretest was statistically significant ($\gamma_{11} = -.01, t = -2.17, df = 16, p = .045$). This means that for every one-point increase in teacher TQM scores, the effect of the pretest on mathematics PLE scores decreases by .01 point. Finally, statistically significant variability in the mathematics PLE means still exists ($\tau_{00} = 1.72, \chi^2 = 336, df = 16, p = .000$), as well as statistically significant variability in the effect of the pretest ($\tau_{11} = .02, \chi^2 = 40.85, df = 16, p = .000$).

**English Descriptives**

An outlier analysis was conducted on all variables in the full model and none were found (i.e., $z > \pm 1.96$). Thus, the final analysis sample contained all 952 students. The descriptive statistics for the Level 1 variable, English pretest, had an average score of 6.33 ($SD = 2.32$). Posttest English scores were lower on average ($M = 3.87, SD = 2.12$). The Level 2 variable (i.e., the TQM) sample again contained 18 teachers with an average TQM score of 87.00 ($SD = 12.54$). The scores ranged from 71 to 112.

**English Assumptions**

Again, normality was tested for all variables (i.e., pretest English, posttest English, as well the TQM). Histograms and skewness and kurtosis statistics suggested that normality was upheld.
English One-Way Random Effects ANOVA

Table 11 below shows the results of One-Way Random Effects ANOVA Model (i.e., the Empty Model). The average student English PLE mean was statistically different from zero ($\gamma_{00} = 4.31$, $t = 10.98$, $df = 17$, $p = .000$). Considerable variation in the student English means still exists ($\tau_{00} = 2.87$, $X^2 = 962.50$, $df = 17$, $p = .000$). The proportion of the total variance in English PLE achievement that can be attributed to the teacher was 54.98%. That is, approximately 55% of the variability in English PLE achievement can be attributed to teacher quality. Based on the significant amount of unexplained variability, additional Level 1 predictors (i.e., pretest scores) were added to try and reduce the variation within students, as well as adding other Level 2 variables to explain between student differences in the following models.

Table 11

One-Way Random Effects ANOVA Model with the English Primary Leaving Examination (PLE)

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient (SE)</th>
<th>$t$ ($df$)</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model for Mean Teacher PLE English Scores ($\beta_0$)</td>
<td>Intercept ($\gamma_{00}$)</td>
<td>4.31 (.39)</td>
<td>10.98 (17)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Variance</th>
<th>$df$</th>
<th>$X^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variation in Teacher Means ($\tau_{00}$)</td>
<td>2.87</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variation within Teachers ($\sigma^2$)</td>
<td>2.35</td>
<td>17</td>
<td>962.50</td>
<td>.000</td>
</tr>
</tbody>
</table>

Note. Deviance (REML) = 3577.74; 2 estimated parameters
**English Unconditional Model**

Table 12 below shows the results of the Unconditional Model with the English pretest as the sole predictor at Level 1 and no Level 2 variables. After including the pretest (i.e., grand-mean centered) as a predictor of English PLE achievement, within-teacher variability in English PLE scores was reduced by 21.28% relative to the One-Way Random Effects ANOVA Model. Overall, English PLE scores across teachers were still significantly different from zero ($\gamma_0 = 4.10, t =$

Table 12

*Unconditional Model with the English Primary Leaving Examination (PLE)*

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient (SE)</th>
<th>t (df)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model for Mean Teacher PLE English Scores ($\beta_0$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept ($\gamma_0$)</td>
<td>4.10 (.36)</td>
<td>11.47 (17)</td>
<td>.000</td>
</tr>
<tr>
<td>Model for Pretest Slopes ($\beta_1$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept ($\gamma_1$)</td>
<td>.36 (.06)</td>
<td>6.37 (17)</td>
<td>.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Variance</th>
<th>df</th>
<th>$X^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variation in Teacher Means ($\tau_{00}$)</td>
<td>2.37</td>
<td>17</td>
<td>637.41</td>
<td>.000</td>
</tr>
<tr>
<td>Variation in Pretest Slopes ($\tau_{11}$)</td>
<td>.05</td>
<td>17</td>
<td>93.71</td>
<td>.000</td>
</tr>
<tr>
<td>Variation within Teachers ($\sigma^2$)</td>
<td>1.85</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Deviance (REML) = 3378.20; 4 estimated parameters
11.47, \(df = 17, p = .000\). On average, across teachers, student pretest scores were positively and statistically significantly related to English PLE achievement within teachers (\(\gamma_{10}\)). The average effect (i.e., slopes) across teachers for the pretest is represented as an increase of .36 points on the English PLE for every one-unit increase in pretest scores.

Statistically significant differences (variability) in the 18 teacher means still exists (\(\tau_{00}\)). There was significant variability in both the intercepts and slopes. Statistically significant variability in the English PLE means still exists after considering the pretest (\(\tau_{00} = 2.37, \chi^2 = 637.41, df = 17, p = .000\)), as well as statistically significant variability in the effect of the pretest (i.e., the slopes) across teachers (\(\tau_{11} = .05, \chi^2 = 93.71, df = 17, p = .000\)). Based on the significant amount of unexplained variability, a Level 2 predictor (i.e., teacher TQM scores) was added to explain the variation between students.

**English Contextual (Full Model)**

After adding the Level 2 predictor (i.e., teacher TQM scores grand-mean centered) to the model, 45.99% of the variance in English PLE achievement can be explained by teacher TQM scores (see Table 13). Additionally, 20% of the variance in the effect of the pretest (i.e., the slopes) can be explained by teacher TQM scores. English PLE scores across teachers was still significantly different from zero (\(\gamma_{00} = 4.10, t = 15.88, df = 16, p = .000\)). That is, the average teacher has mean English PLE scores of 4.10 for their students. Teacher TQM score had a statistically significant influence on the intercept for English PLE achievement. The effect of teacher TQM scores on mean PLE English achievement is negative and statistically significant (\(\gamma_{01} = -.09, t = -4.20, df = 16, p = .000\)).
Table 13

(Contextual (Full) Model with the English Primary Leaving Examination (PLE)

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient (SE)</th>
<th>t (df)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model for Mean Teacher PLE English Scores (β₀)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept (γ₀₀)</td>
<td>4.10 (.26)</td>
<td>15.88 (16)</td>
<td>.000</td>
</tr>
<tr>
<td>Teacher Quality Measure (TQM; γ₀₁)</td>
<td>-0.09 (.02)</td>
<td>-4.20 (16)</td>
<td>.001</td>
</tr>
<tr>
<td>Model for Pretest Slopes (β₁)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept (γ₁₀)</td>
<td>.36 (.06)</td>
<td>7.13 (16)</td>
<td>.000</td>
</tr>
<tr>
<td>Teacher Quality Measure (TQM; γ₁₁)</td>
<td>-0.01 (.00)</td>
<td>-2.31 (16)</td>
<td>.035</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Variance</th>
<th>df</th>
<th>X²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variation in Teacher Means (τ₀₀)</td>
<td>1.28</td>
<td>16</td>
<td>379.78</td>
<td>.000</td>
</tr>
<tr>
<td>Variation in Pretest Slopes (τ₁₁)</td>
<td>0.04</td>
<td>16</td>
<td>83.38</td>
<td>.000</td>
</tr>
<tr>
<td>Variation within Teachers (σ²)</td>
<td>1.85</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Deviance (REML) = 3381.50; 4 estimated parameters

*p = .001.* The value of -0.09 can be interpreted as the decrease in a teacher’s mean English PLE achievement to be expected for a one-unit increase in teacher TQM scores, on average. The average teacher, therefore, is predicted to have a mean English PLE achievement, on average, of 4.10 for their students, and with every one-point increase in TQM scores (i.e., increasing teacher quality), the mean English PLE scores would decrease by .09 points. Thus, more quality teachers have students with better English PLE scores on average.

Again, the effect of the pretest on English PLE achievement is positive on average and statistically significant (γ₁₀ = .36, t = 7.13, df = 16, p = .000). Thus, there was a
statistically significant effect of the pretest slope (i.e., predicting the English PLE scores) across students and teachers. Specific to the research question, the impact of the TQM on English PLE scores controlling for the pretest was statistically significant ($\gamma_{11} = -.01, t = -2.31, df = 16, p = .035$). This means that for every one-point increase in teacher TQM scores, the effect of the pretest on English PLE scores decreases by .01 point. Finally, statistically significant variability in the English PLE means still exists ($\tau_{00} = 1.28, \chi^2 = 379.78, df = 16, p = .000$), as well as statistically significant variability in the effect of the pretest ($\tau_{11} = .04, \chi^2 = 83.38, df = 16, p = .000$).

**Research Question 2C**

Research Question 2C (i.e., “What is the relationship between school type and a pupil’s achievement on the PLE (i.e., 7th grade) in Uganda?”) was analyzed using a One-Factor Analysis of Covariance (ANCOVA). This was conducted to determine the relationship between a pupil’s school type (i.e., urban or rural, essentially school SES) on the PLEs in Uganda, while controlling for pupils’ prior mathematics and English achievement/knowledge.

**Mathematics Descriptives**

Before conducting the analysis, outliers were removed ($n = 49$). That is, all PLE pretest and posttest scores with $z$ scores $>\pm 1.96$ for were removed. Thus, the final analysis sample contained 678 urban and 225 rural ($N = 903$). Overall, the pretest mean was higher ($M = 7.05, SD = 2.09$) than the posttest mean ($M = 4.88, SD = 2.31$). For the pretest, the urban schools had higher scores ($M = 6.77, SD = 2.10$) compared to the rural
schools ($M = 7.91, SD = 1.79$). Additionally, for the posttest, urban schools’ scores were lower ($M = 4.93, SD = 2.45$) than the rural schools ($M = 4.72, SD = 1.84$).

**Mathematics Assumptions**

Independence was tested via an examination of residuals. A random display of points of the residuals against values of independent variable (i.e., urban and rural) provided evidence of independence. Normality was tested on the covariate (i.e., pretest mathematics), the dependent variable (i.e., posttest mathematics), and separately in each group (urban and rural). Again, the assumption was tested via examinations of S-W, skewness and kurtosis, histograms, and Q-Q plots.

Calculated $z$ scores for skewness and kurtosis indicated that there was some significant skewness and kurtosis, but not for all (i.e., pretest mathematics ($S = -7.83, K = -6.25$), posttest mathematics ($S = 1.93, K = -6.54$), urban pretest ($S = -4.40, K = -6.66$), rural pretest ($S = -9.79, K = -4.29$), urban post ($S = .90, K = -6.88$), and rural posttest ($S = -.74, K = .86$). The S-W test was also significant ($p < .05$). Histograms and Q-Q plots suggested relatively normal distributional shapes across all the variables and in the groups. Overall, there is evidence that normality has been met.

Linearity of the dependent variable with the covariate was examined with scatterplots, both overall and by group. Overall, the scatterplots suggested positive linear relationships. Homogeneity of Regression Slopes was suggested by similar regression lines evidenced in the scatterplots of the dependent variable and covariate by the independent variable (reported earlier as evidence for linearity). There was a significant
interaction between the main effect and the pretest \((F (1, 899) = 12.85, p = .000)\). The Homogeneity of Variance assumption was not satisfied \((p < .05)\).

**Mathematics ANCOVA**

A One-Factor ANCOVA was conducted to assess the effect of School Type (i.e., urban or rural) on student mathematics achievement controlling for mathematics pretest. The IV was School Type. The DV was scores on the PLE mathematics exam (i.e., the posttest) that is administered following completion of the 7th grade. Scores on a mock mathematics test administered at the beginning of the 7th grade were used as the covariate (i.e., the pretest). The results of the ANCOVA suggest a statistically significant main effect of the covariate on the DV \((F = 408.724, df = 1, 900, p < .001)\). There were also statistically significant main effects of School Type \((F = 40.788, df = 1, 900, p < .001)\) on the posttest adjusting for the pretest. The effect size for School Type was small \((\text{partial } \eta^2 = .043)\) and large \((\text{partial } \eta^2 = .312)\) for the covariate. Observed power was strong for both.

The results indicate that there are significant differences between urban and rural schools on mathematics achievement controlling for prior mathematics ability. Specifically, the adjusted means indicate that rural schools have higher mathematics achievement on average \((M = 4.16, SE = .13)\) compared to urban schools \((M = 5.12, SE = .07)\). Note that the lower mean indicates better performance, and a higher mean indicates poorer performance on the PLE exams. (i.e., 1 is the highest score and 9 is the lowest score).
**English Descriptives**

Before conducting the analysis, outliers were removed ($n = 100$). That is, all PLE pretest and posttest scores with $z$ scores $\pm 1.96$ for were removed. Thus, the final analysis sample contained 662 urban and 190 rural ($N = 852$). Overall, the pretest mean was higher than the posttest mean ($M = 6.30$, $SD = 2.20$ and $M = 3.45$, $SD = 1.58$, respectively). For the pretest English scores, the urban schools had higher scores ($M = 6.20$, $SD = 2.20$) compared to the rural schools ($M = 6.64$, $SD = 2.16$). Additionally, for the posttest, urban schools’ scores were higher ($M = 3.40$, $SD = 1.50$) than the rural schools ($M = 3.62$, $SD = 1.85$).

**English Assumptions**

Again, the assumption of Independence was tested via the scatterplot of residuals, which confirmed that the assumption was met. Normality was tested on the covariate (i.e., pretest English scores), the dependent variable (i.e., posttest English scores), and separately in each group (urban and rural). The assumption was tested via examinations of S-W, skewness and kurtosis, histograms, and Q-Q plots.

$Z$ scores for skewness and kurtosis were calculated. It was found that pretest English ($S = -3.61$, $K = -6.46$) posttest English ($S = 7.58$, $K = -3.48$), urban pretest ($S = -2.85$, $K = -5.71$), rural pretest ($S = -2.38$, $K = -2.99$), and urban posttest ($S = 6.11$, $K = -2.45$) and rural posttest ($S = 3.63$, $K = 3.86$) were mostly significantly skewed and kurtotic. S-W test for normality was significant across pretest and posttest and in each group ($p < .05$). However, histograms and Q-Q plots suggested relatively normal
distributional shapes across the variables and in each group. Overall, there is evidence that normality has been met.

Linearity of the dependent variable with the covariate was examined with scatterplots, both overall and by group. Overall, the scatterplots suggested a positive linear relationship. Homogeneity of Regression Slopes was examined by similar regression lines evidenced in the scatterplots of the dependent variable and covariate by the independent variable (reported earlier as evidence for linearity). This assumption was confirmed by a nonstatistically significant interaction \((F(1,848) = 2.77, p = .097)\). The Homogeneity of Variance was not satisfied \((p < .05)\).

**English ANCOVA**

A One-Factor ANCOVA was conducted to assess the effect of School Type (i.e., urban or rural) on student English achievement controlling for English pretest. The IV was School Type. The DV was scores on the PLE English exam (i.e., the posttest) that is administered following completion of the 7\(^{th}\) grade. Scores on a mock English test administered at the beginning of the 7\(^{th}\) grade were used as the covariate (i.e., the pretest). The results of the ANCOVA suggest a statistically significant main effect of the covariate on the DV \((F = 344.56, df = 1, 849, p < .001)\). There was a nonsignificant main effect of School Type \((F = .143, df = 1, 849, p = .706)\) on the posttest adjusting for the pretest. There was a large effect size for the covariate \((\text{partial } \eta^2 = .289)\) and observed power was strong.

The results indicated that in English, on average, urban schools scored only .04 points higher than rural schools according to the adjusted means. This might explain why
there were no significant differences between urban and rural schools on English achievement controlling for prior English ability. The adjusted means indicated that urban schools have higher English achievement on average ($M = 3.44, SE = .05$) compared to rural schools ($M = 3.48, SE = .10$). Note that the lower mean indicates better performance, and a higher mean indicates poorer performance on the PLE exams. (i.e., 1 is the highest score and 9 is the lowest score).
CHAPTER 5

DISCUSSION

Research Question 1

The first research question examined the psychometric properties of the newly developed TQM. The measure containing the original 80 items, after reviewer feedback, was examined for Likert category functioning along with eliminating misfitting items. It was discovered that combining the adjacent categories 0 and 1 was necessary, which means that the Likert category structure went from a 5-point scale (i.e., 0 = Never, 1 = Few, 2 = Many, 3 = Most, 4 = Almost All) to a 4-point scale (i.e., 0 = Few, 1 = Many, 2 = Most, and 3 = Almost All). Collapsing those categories improved Likert category definition in that the average measures and step calibrations became monotonic, and the probability curves showed that each category represented a distinct portion of the underlying variable (Bond & Fox, 2007). Misfitting items were also examined. Twenty-six items were eliminated based on low correlations with the rest of the measure (i.e., similar to item-total correlation in reliability analysis), high infit and outfit values, and redundancy with other items due to placement on the vertical ruler.

After elimination of the above 26 items, the 54-question TQM produced better psychometric properties than the original 80-question version (e.g., Coefficient Alpha = .96). With the exception of one item (i.e., 43), point-measure correlations for this model were high. Infit and outfit diagnostics were above 2.0 for nine items (i.e., unstandardized
or standardized), which is considered to be poorly fitting. Additionally, most of the persons were above the item mean on the vertical ruler. The items covered a larger logit range in difficulty compared to persons. Even though adding “difficult” items was one option, these items were instead deleted. Moreover, at several points on the scale, there were items at the same position on the vertical ruler, which indicates that these items were redundant. Therefore, based on these concerns presented above, a more detailed content analysis of some of the items was implemented. This was also needed as a 54-question survey was deemed too long to be completed by teachers.

From the 54-item TQM, 16 more items were removed based on content analysis and a general reliability analysis. This produced the 38-item TQM with acceptable psychometric properties, and is also a more portable and efficient measure compared to the earlier versions. In the 38-item TQM, again, most of the persons appeared above the item mean on the vertical ruler. This is to be expected on such a measure that contains questions in which teachers would want to portray themselves in a more favorable way (i.e., hence the higher scores and a bias towards the higher end of the Likert scale). The items covered a narrower range in difficulty compared to persons. As mentioned previously, this indicates that more “difficult” items may need to be added in the future. Difficulty in this instance means adding items pertaining to teacher quality that some teachers would not be able to highly endorse (i.e., endorse the higher Likert categories). Also, at several points there were still items at the same position on the vertical ruler, which indicated that there were still items that may be redundant.
The 38-Item TQM

The 38-item TQM produced acceptable psychometric properties (e.g., Coefficient $\alpha = .88$). The main concern of this final version of the TQM is that most of the persons appeared above the mean on the vertical ruler, and the items covered a narrower range compared to the persons. This indicates that more “difficult” items may need to be added in the future. Also, at some points on the scale there were items at the same position on the vertical ruler, which indicated that these items may be redundant. The concerns presented above regarding this version of the TQM warrant further investigation of the construct and this measure.

Research Question 2A

This research question examined the relationship between a pupil’s sex and achievement on the PLE (i.e., 7th grade) in Uganda using a Two-Factor ANCOVA. For this research question, the discussion of the results will be separated into two parts – sex and SES. The first part will contain a discussion of the sex and SES differences in relation to pupil mathematics achievement. The second part will contain a discussion of sex and SES differences and pupil English achievement (although SES was not the primary variable of interest in this research question).

Sex and Mathematics Performance

For mathematics performance, regarding sex differences, the current study found that there are statistically significant differences in mathematics achievement on the PLE (i.e., after controlling for prior achievement/knowledge) between males and females, with males having higher scores. These results concur with the findings of prior studies that
show that males outperform females in mathematics (UNESCO, 2012). The results also conform to one Ugandan study that showed that in mathematics, boys perform better than girls (Kasente, Nakanyike, & Balihuta, 2003).

Additionally, girls’ intelligence is underrated in the society. According to Dweck (1986), ability in mathematics is thought to be hereditary and external factors cannot change it. Girls, as a result, are demoralized and do not see themselves capable to perform at the same level as boys.

This suggests that when girls do not perform as expected, it is because of limited ability. Consequently, the society indirectly encourages boys to work harder, and not girls. This, coupled with the belief that girls are “inferior” to boys in Uganda may have a far reaching effect on girls’ learning outcomes. Negative social attitudes, as well as cultural practices may, therefore, explain why boys outperform girls.

Another potential explanation for the differences includes research by Graven (2012). The author contends that achievement differences due to sex in developing countries like Uganda are more pronounced because of a “third variable.” That is, there is a correlation with gender-based tasks that hamper school preparation. For instance, if girls have more domestic chores at the beginning and the end of each day, they will have less time to complete homework. For instance, these might include walking long distances to collect water or firewood, cooking, cleaning, and taking care of the younger siblings and elderly family members. Consequently, they may even be forced to miss days of school in order to focus on domestic tasks. This may contribute to the differential
mathematics performance between boys and girls. Additionally, it may be an explanation to the statistically significant results in this current study.

**SES and Mathematics Achievement**

Further examination of Research Question 2A found that there were statistically significant differences in achievement on the PLE between boarding schools and day schools. The results indicated that boarding schools (i.e., a proxy for higher student SES) have a higher achievement level than their day counterparts in mathematics. This finding is consistent with the conclusions of Nanyonjo (2007) that pupils who live further from school (day pupils) perform worse than those who live near the school (boarding pupils). The finding in the current study is not surprising, as pupils who live further away from school are likely to arrive late at school, a factor that may lead to poor performance. That is, these pupils spend more time walking to and from school and thus reducing the time available for study. Moreover, they are also likely to be more exhausted as a result of the long traveling and therefore their level of concentration may be reduced.

Again, these findings are not surprising because while the Universal Primary Education (UPE) program has increased access to primary education, especially among the pupils with low SES, it can be seen that other costs of primary education, such as school uniforms, continue to hinder the educational attainment of many children in day schools. In addition, the provision of quality education remains a challenge. This is consistent with findings in a study by Uwezo (2010). The study concluded that there are disappointing levels of learning outcomes among primary school children from poor households. Additionally, the continued and consistent dominance of boarding schools,
which have pupils from richer households in the Primary Leaving Examinations (PLE), is yet another explanation of the significant results in the current study.

The researcher believes that the findings are valid because day school parents may not be able to settle their school dues by a set date. Pupils from middle or high social economic background are able to abide by this requirement and settle down to study whereas those from poor socio-economic backgrounds may not find the fees easily and this may affect their performance at university. This could be the reason for the significant results in both mathematics and English.

**Sex and English Performance**

The results found no statistically significant differences in English achievement on the PLE (i.e., after controlling for prior achievement/knowledge) between males and females. This finding is interesting in that earlier research has demonstrated that girls perform better than boys with regard to English performance and other related measures (e.g., reading acquisition).

Past studies show girls are more likely to performance better than boys in English because girls tend to use their free time in reading and communication as compared to the boy (Zuze & Leibbrandt, 2011). However, the nonsignificant difference does corroborate evidence from one study (Nanyonjo, 2007) that found no significant difference between the Ugandan boys and girls in English. To explain these results, it may be important to consider some of the performance interventions that have been implemented alongside UPE. In recent years, school districts have attempted to bridge the gender gap, which has contributed to boys beginning to outperform girls in English. Additionally, existing social
relations of sex that positions girls as providers of homestead labor may also explain the mixed results. As seen in the literature, the girls are socialized as apprentices of their mothers so that they can play similar roles in future. This factor may be critical in explaining why girls have not performed as hypothesized.

The adaption of the UPE saw massive numbers of pupils enrolled, but without commensurate expansion of facilities, like increased number of qualified teachers, adequate teaching, and learning materials. Therefore, the quality of education could have been compromised. Moreover, it can also be seen that under UPE, parents are supposed to provide lunch, exercise books, uniform, pens and ensure hygienic conditions, especially of the girls. That is, boys are easily allowed to work for money to spend on school needs in Uganda, girls are generally confined within their home chores. Consequently, boys have easier access to money for lunch and learning materials unlike girls. Therefore, there is gender inequality in the allocation of household resources, since girls have access to less time to attend school as well as work for money. Such differences in opportunities may contribute to a decline in learning outcomes of girls in English.

From another perspective, girls usually require extra financial support for personal hygiene as compared to boys. Due to the meager income in the Ugandan society, girls have not been provided with items like sanitary pads, among others. Consequently, they decide to not attend class. Kasante, Nakanyike and Balihuta (2003) show that such situations will affect the learning outcomes of girls, and places boys at an advantage. This
finding explains the nonsignificant result in the current study, and may as well be the reason the girls did not perform better than boys in English as had been hypothesized.

**SES and English Performance**

The analysis for this research question found that there were statistically significant differences in achievement on the PLE between boarding schools and day schools. The results indicated that boarding schools (i.e., a proxy for higher student SES) have a higher achievement level than their day counterparts. The explanations for these findings are similar to the explanations for mathematics performance. As explained above, socio-economic factors do explain the disparity in performance between boarding pupils and day pupils. Also, cultural constraints negatively impact achievement level among students. Additionally, it was described that day children who walk long distances to school, as a result of their socio-economic environment, lagged behind in learning outcomes.

The findings of this study confirmed the findings of Heyneman and Loxley (1982; 1983) who concluded that, in virtually all nations, children of parents high on the educational, occupational, and social scale have far better chance of future educational success compared to children of day laborers or farmers. Heyneman and Loxley (1982; 1983) add that the findings of many empirical studies suggest that children whose parents are at the bottom of the socio-economic hierarchy are not as inclined to seek or gain access to available educational facilities as the children with families located at the middle or top of the hierarchy, which can contribute to differences in learning outcomes.
Research Question 2B

This research question (i.e., “What is the relationship between teacher quality and a pupil’s achievement on the PLE (i.e., 7th grade) in Uganda?”) was analyzed using a Two-Level Hierarchical Linear Model (HLM). HLM was used to investigate the relationship between teacher quality and pupils’ achievement on PLE achievement in mathematics and English controlling for prior achievement/knowledge. Student scores in English and mathematics on the PLE, as well as the survey scores from the newly developed TQM were used. The study hypothesized that TQM scores would be a significant predictor of PLE scores in mathematics and English with higher TQM scores predicting higher PLE scores (i.e., a positive relationship).

The newly created TQM produced teachers’ average score across the four domains: (1) “Planning and Preparation,” which incorporates components pertaining to the knowledge and skills needed to plan and orchestrate a day of learning in the life of a child, (2) “Classroom Environment,” which covers crucial components such as providing a physically safe environment for all students, (3) “Instruction,” which includes items that evaluate teachers’ ability to engage students in active learning that promotes critical thinking, decision-making, and problem solving while using a variety of instructional strategies, and (4) “Professional Responsibilities,” which contains items critical to teacher education programs such as demonstrating a high level of professionalism in their work with children, families, colleagues, and the school community.

The results indicated that with every one-point increase in TQM scores (i.e., increasing teacher quality), the mean student mathematics PLE scores decrease by .01
point. Thus, more quality teachers have students with better mathematics PLE scores on average. Similarly, with every one point increase in TQM scores (i.e., increasing teacher quality), the mean student English PLE scores also decrease by .01 point. Thus, more quality teachers have students with better English scores. It is apparent that although teacher quality is a predictor of increased student achievement, it is only to a small extent. That is, student scores only improved by .01 points on the PLEs. However, this uniform change across subjects is interesting to note in that this may provide evidence of the utility of the TQM as an indicator of student English and mathematics performance.

Generally, the HLM results showed that indeed teacher quality, as defined by the four domains in the newly created measure, is a significant predictor of pupil performance in both mathematics and English. This finding is consistent with the findings in the majority of studies that concluded that teachers are one of the most critical factors that predict variation in student test scores (Abbott & Ryan, 2001; Battistich & Hom, 1997; Goldhaber & Brewer, 1997; Nanyonjo, 2007). Moreover, Sanders and Rivers (1996) found that students assigned to ineffective teachers over the course of several years demonstrate significantly lower academic achievement than those pupils that are assigned to high quality teachers.

**Research Question 2C**

This research question (i.e., “What is the relationship between school type and a pupil’s achievement on the PLE (i.e., 7th grade) in Uganda?) was analyzed using a One-Factor ANCOVA. The relationship between school type and pupils’ achievement on PLEs in mathematics and English controlling for prior achievement/knowledge was
investigated. Student scores in English and mathematics on the PLE, as well as the school type (i.e., urban or rural) were used. Rural schools were used as a proxy for lower school SES, whereas urban schools were a proxy for higher school SES.

**School SES and Mathematics Achievement**

The findings showed that there is a statistically significant difference between school types (i.e., urban versus rural) on pupil achievement in mathematics. That is, the rural school (low SES) had higher mathematics scores compared to the urban (high SES) schools. This result is surprising as numerous studies have shown that school SES has a link with academic achievement, with higher SES schools having higher achievement in general (Caldas & Bankston, 1997; Rumberger & Palardy, 2005; Zhang, 2006). Additionally, the only study done in Africa, SACMEQ II, found that in many less developed countries, education in rural areas is often synonymous with a poor context for learning. That is, primary schools in urban areas of Sub-Saharan Africa consistently outperform their rural counterparts by substantial margins.

Although no data in Uganda was found to compare with the findings in this study, examining the research literature in the United States, urban schools are at a disadvantage when viewing results for students in the area of mathematics. However, in the U.S., it is not a surprise as the urban schools are the ones with the lower SES. Results show that a large percentage of U.S. students attending urban middle schools achieve low levels of mathematics proficiency (Ruby, 2006). Ruby notes several issues which apply to Uganda as well. Lack of learning material coupled with insufficient funding is among the critical problems. In Uganda, it can be seen that, since the introduction of the UPE, urban schools
no longer get funding from the parents as they used to do. The meager UPE resources cannot sufficiently cover their needs. Moreover, considering higher standard of living in urban areas, these urban schools tend to be at a greater disadvantage compared to the rural schools. Since lack of materials is a determinant of student achievement (Nanyonjo, 2007), the significant results found in this study, may be partly explained by insufficient resources.

School SES and English Achievement

This sub-section of Research Question 2C examined the relationship between school type and pupils’ achievement on PLEs in English controlling for prior achievement/knowledge was investigated. As explained previously, student scores in English on the PLE, as well as the school type (i.e., urban or rural) were used. The findings showed that there are no statistically significant differences between school types on pupil achievement in English. This result is surprising in that numerous studies have shown that school SES has a link with academic achievement, with higher SES schools having higher achievement in general (Caldas & Bankston, 1997; Rumberger and Palardy, 2005; Zhang, 2006).

There are a number of important methodological limitations in this study that may explain these results. For example, the relatively small sample size of rural schools (225) in comparison to relatively large sample of urban schools (678) could affect the validity of the conclusions to be drawn in this finding. Specifically, unequal sample sizes can be problematic with violated assumptions as was the case with the data at hand (Lomax &
Hahs-Vaughn, 2012). For example, with Homogeneity of Regression slopes, if a violation occurs, the slope can yield biased adjusted means and can affect the \( F \) test.

Additionally, when subjects are not assigned to the treatment groups at random, interpreting the analysis of covariance can be particularly troublesome (Lomax & Hahs-Vaughn, 2012). The results of this study need, therefore to be interpreted with care as the subjects were not assigned randomly, and the study does not use a true experimental design. Lomax and Hahs-Vaughn (2012) suggest that the ideal application for an analysis of covariance is an experiment in which subjects are randomly assigned to treatments. Thus, the expected value of the covariate mean for each group and any differences can be attributed only to chance. Moreover, when ANCOVA is used this way, the error term is significantly reduced (Lomax & Hahs-Vaughn, 2012).

**Implications**

**Research Question 1**

The results from this research question have implications for the measurement of teacher quality in Uganda, and also for administrators and school districts in Uganda looking for quick and efficient ways to gauge general teacher quality and diagnose weaknesses. The analyses revealed that there are several areas that comprise the construct of teacher quality in Uganda, and the domains of teacher quality realized through qualitative data analysis appear similar to domains outlined by Danielson (1996). Rasch Analysis indicated that the 38-item TQM was an acceptable beginning in producing a psychometrically sound measure of the construct.
Although refinement of the measure (e.g., eliminating misfitting items, obtaining a larger sample size and conducting a factor analysis, validation through various mechanisms such as corroborating self-reported information with administrator, peer, or outsider observation, etc.) is needed, and more research on the construct needs to be done, the results from the initial development of the TQM support its eventual use as a way to gauge teacher quality or effectiveness in general in Uganda. For school districts, schools, administrators, and teachers, the measure is intended to be used as a diagnostic tool for determining teaching staff needs in the school system. With this measure, some areas of weakness can be identified, and can highlight where teachers may need remediation and further training.

The demand for school systems, individual schools, and teachers to be accountable has increased considerably in the past two decades in the United States. It is becoming an important issue to address in other countries such as Uganda. This demand for accountability relates to a measurement of attainment of educational standards and objectives. Specific performance measures of those standards for students and teachers alike that track competency gains are a requirement for most educational systems. Thus, this measure of teacher quality was created to attempt to accurately define and measure the construct and assist in further gauging attainment of educational standards and objectives. Use of this measure in Uganda and other African countries (in the future pending further validation) may begin to provide information on teacher quality that could inform policy and professional development.
Research Question 2A

**Sex.** There were significant differences between boys and girls on mathematics, but not English performance, with boys outperforming girls in mathematics. Senior decision-makers of the 15 Ministries of Education in Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) have been concerned with the gender equality issue not only in access and participation, but also in learning achievement (Saito, 2011). In Uganda, the gender gap (dominated by boys) has even widened in recent years in mathematics (Saito, 2011). Thus, these results have particular significance for international organizations, donors, and the government of Uganda that have expended enormous resources and efforts on educational programs that have aimed to reduce gender differences in education. These results seem to suggest the need to focus sex-related interventions on mathematics achievement for girls, instead of focusing on “access” and “participation.” Additionally, there is a need to have in place a community adult education, as well as media to promote parents interest in the education in the education of their children. Additionally. The Ministry of Education should make themes of gender equality part of the core curriculum.

**SES.** There were differences in student SES (i.e., boarding and day schools) on mathematics and English achievement, with boarding (i.e., higher student SES) school students performing better than their day school student counterparts. Based on these significant findings, there are strategies to help reduce the effects of SES on pupil achievement. For example, there is some evidence from SACMEQ II data that most low SES parents in Uganda show little or no interest in their children's school work. Even
though, changing the attitudes of parents is as crucial for challenging such gender stereotype. A campaign to sensitize parents through media and communities can go a long way to change the way parents perceive educating their children especially in rural day schools. Moreover, there is need for equal distribution of qualified teachers across in boarding schools as well as in day schools.

**Research Question 2B**

In the investigation of the relationship between teacher quality, and pupils’ achievement on PLE achievement in mathematics and English controlling for prior achievement/knowledge, the results show that there is significant relationship. The findings in this research question imply that it is possible to see a relationship between student achievement and teacher quality. Overall, these findings have implications for schools that may use the newly created TQM to identify quality teachers. Identifying quality teachers and more importantly, where there are deficits based on the four domains, can assist school districts in staging interventions to improve teacher performance. That is, the findings are a guide to the MoES to identify teachers with performance deficits so as to possibly plan a teacher development program for them.

Adding the TQM to a school’s accountability system would provide a critical empirical perspective to the multifaceted process of teacher evaluation, which is lacking in Uganda. Also, diagnosing weaknesses, and providing professional development opportunities, can increase the educational success of more students (specifically in mathematics and English, from this study). Studies such as this may help the Ugandan MoES to better understand the link between teacher quality and desirable student
outcomes. Moreover, by focusing on the domains of quality teaching outlined in this study, teacher training institutions may be better equipped to educate teachers on the importance of teacher quality related to student performance, and to set meaningful performance expectations once teachers are in classrooms.

**Research Question 2C**

**School type.** Examining the relationship between school type and pupils’ achievement on the PLEs, the results showed that there were no differences in mean English scores for urban and rural schools. However, there were significant differences between urban and rural schools on student mathematics scores, with rural schools having higher mathematics achievement. This was surprising in that rural schools tend to be particularly vulnerable to educational disadvantage compared to their urban school counterparts.

As explained before, even though socioeconomic factors were addressed in the reviewed literature, the amount of research completed on these factors is limited in Uganda. The importance of pupils’ learning resources is still overlooked in many of the studies, as well as the effects that these variables have on student achievement. Specifically, there seems to be no literature comparing 7th grade in rural and urban schools in the Primary Leaving Examinations.

The results in this study are therefore among the first ones, and so it is critical to consider designing future studies on specifically 7th grade to validate the current findings.

There were non-significant differences found between rural and urban schools on English performance. The government could consider allocating more UPE funds to rural
schools for the purchase of basic learning materials and textbooks, especially those schools serving poor communities, as well urban schools in slum areas. Alternatively, the government may wish to consider introducing or encouraging charity or donor programs for providing pupils from poor backgrounds with basic learning materials and textbooks, irrespective of whether these pupils are in rural or urban schools.

**Research Question 2 (ALL)**

While contributing to the body of knowledge on education in Uganda, these research findings will be particularly useful to the Ugandan MoES and the proprietors involved in teaching services. The results can serve as a guide in designing interventions capable of improving the quality of teachers. For instance, the content of the newly created measure should be included in the teacher development module to supplement the packages used currently. Studies have found that the teachers who are trained under the Teacher Development Management System (TDMS) did not show any kind of performance, suggesting that they are lacking in content.

The results from this research question were examined within the framework of UPE, and they may contribute to policy and practical implementation of the UPE program, as the teacher quality variable had not been given due important. The focus on the pupil, teacher, and school (i.e., a multi-level approach) can provide insight into the many influences on pupil academic performance under UPE and provide suggestions based on empirical results. Therefore, the findings of this study will provide a vital resource to a broad audience of educators, policy makers, and funding partners.
Finally, this study provides a timely contribution to the body of knowledge on educational reforms in Uganda since multi-level studies on the factors that influence pupil performance in English and mathematics on the PLEs are not available at this time. Most studies address enrollment issues, pupil-teacher ratios, and other factors that influence student performance (Nanyonjo, 2007). However, these studies have not investigated the relationship between the multi-level factors that influence pupil academic performance on high-stakes exams such as the PLEs in mathematics and English, especially the influence of the quality of the teacher.

**Limitations and Future Directions**

**Research Question 1**

The newly-created TQM is a behavior frequency measure. There are a number of limitations with such self-reporting instruments, specifically with regard to validity and reliability. Participants may produce inaccurate responses so as to suit their particular self-presentation. The teachers may present themselves in a socially desirable way and, therefore, provide responses that make them appear more acceptable to the researcher (DeVellis, 2003). Although there are practical benefits to using self-report measures, this mode of assessment is subject to social desirability, with the TQM being no different. The possibility of bias and distortion on the part of the subjects in light of their own motives, self-interest, or desire to “look good” is elevated with this type of measure. This is also related to the obtrusiveness of the measure. It is highly apparent what construct or trait the TQM is measuring, and it is likely that most teachers responded to most of the items with the highest category (i.e., Almost All). Although the directions indicated that
these results would not be shared with anyone specifically and that they would be reported collectively, teachers may have endorsed the items in the direction of the scale due to social desirability. Future studies may consider carrying out direct observation of teachers in their classrooms.

Second, there could have been chance of researcher bias. This means that the research assistant might, without intending, administer the survey to different teachers differently (e.g., might emphasize certain parts differently when reading instructions). This may result in small variations in survey administration that can have effects on how participants respond to the items. Although this was minimized due to the fact that the same research assistant administered the survey to all the teachers, this was still a possibility. Future studies may need to use a triangulation approach by collecting comparative data from school administrators, as well. Alternatively, future studies should consider using observational instruments as they can be valid and reliable (Strong, 2011).

Another limitation is the small sample sized used for the Rasch Analysis of the TQM. Approximately 50 to 100 would be ideal are if item calibrations are to be stable within + 1 logits (i.e., 99% CI - 50 people) or + 1/2 logits (i.e., 95% CI - 100 people; Linacre, 1994). Thirty-six teachers were in the final analysis sample, which is just under the ideal amount to have the item calibrations stable within + 1 logits. Future studies not restricted by resources should consider obtaining a larger sample size, representative of teachers across Uganda and all types of schools.

Even though using reliability analysis and Rasch Analysis was helpful in examining and establishing sound psychometric properties of the TQM, the sample size
was too small to conduct EFA. If a larger sample size is obtained, then Exploratory Factor Analysis (EFA) would be very appropriate to use in future studies on the TQM. According to Tabachnick and Fidell (2001), there should be at least 300 cases for factor analysis. The large sample size is critical so as to have reliably estimated correlations. In the current study, however, the larger sample was not obtained due to practical considerations such as time, money, and distance.

It is of concern to note that this study only covered the basic applications of Rasch modeling in measurement development. Several very prominent aspects of Rasch modeling were not considered in this study. Rasch modeling, as presented here, focused on the item and person differentiation, but another aspect of fitting the Rasch model is the ability to check for discrimination of items by demographic variables. In this case, the software can determine if some questions have different difficulties for boys or girls (i.e., Differential Item Functioning [DIF]). If a question is significantly different by gender, then the groups are not measured “on the same scale,” and the results would need to be analyzed separately. Future research may include DIF analysis to examine how items may function based on specific demographic characteristics.

The Rasch model is termed a "strong" model since its assumptions are more difficult to meet than those of Classical Test Theory. The benefits derived from its use come at the cost of, on occasion, failure to define a measure at all. When data do not adequately fit the model, the instrument construction process must begin anew. An overall failure could occur if items are poorly constructed or are not comprehensible to the population, if there is a mismatch between the respondent group's abilities/attitudes
and item difficulties, or there are anomalies in the item-person interactions. The TQM in its present form does not appear problematic with regards to the points mentioned above. However, the measure may be improved using the knowledge from the runs of the data and finding that a separate, abbreviate assessment instrument may exist (i.e., a smaller version than the 38-item TQM). Future research should focus on clearly defining the construct, validating the scores on the revised measure, and eventually demonstrating the potential flexibility of the measure with other content areas such as Social Studies and Science subjects.

**Research Question 2 (All)**

Some other threats to internal validity to consider as limitations in the current research question are testing. Testing refers to the effects that taking a test one time may have on subsequent performance on the test (Cook & Campbell, 1979). That is, changes demonstrated on PLE after the administration of the mock examination (i.e., pretest mathematics and English) may only be attributed to repeated testing. Testing may not be an issue in the current research question as PLE and the pretests were administered seven months apart. Moreover, the items used at each of the administration were not the same. As mentioned previously, the implementation of a no-treatment or control group can great assist in the inferences drawn from the results in this first research question.

Concerning the general threats to internal validity, or the extent to which the intervention rather than extraneous influences can account for the results, some more limitations are apparent (Cook & Campbell, 1979). For example, history could have impacted the results of the current research question if unplanned events may have
disrupted the administration of the PLEs, or even the administration of the TQM. Additionally, maturation may be a limitation in this second research question since the data sampled took place over a few months. Simply growing older and wiser across the months of interest may explain the differences in this research question. Thus, future studies should include a no-treatment or control group to identify the impact of the “intervention.”

Another limitation is that the proposed study focuses on one single grade level (i.e., 7th grade) for two subject areas only (i.e., English and mathematics). There is an external validity concern as the study may not generalize beyond grade seven, as well as to subjects other than the ones in the study. Also, there is an issue with external validity as the investigation is limited schools in the Wakiso District. Moreover, this is a correlational study, and cannot claim a cause and effect relationship. That is, while the term effect may be used to describe the statistical relationship between variables, a true causal relationship cannot be established. Future studies need to consider sampling a number of districts with multiple primary and secondary schools could have added to the validity of the results. Future studies should include other core subjects like Social Studies and Science. Additionally, future studies should consider designing an experimental study with a control group so as to control group so as to claim a causal relationship.

Even though studies show that teacher quality is an important determinant of student achievement (Strong, 2011), measured quality may explain only a small portion of the observed differences in student outcomes across teachers. The reasons for this are
measurement-related (i.e., defining the construct of teacher quality is no easy task), and definitions vary from researcher to researcher, from one community to another community, and from one generation to the next. Therefore, the results of this study on the effects of teacher quality on pupils’ achievement may not be generalizable to all teachers, places, and to all points in time. Future studies may want to investigate other variables that may predict student achievement.

An additional condition that limits the generalizability of the study is the fact that the DV in the study was measured with one single test, the PLE. It would be preferable to measure mathematics and English achievement with gain scores across time using several instruments. Thus, the findings of this study may not generalize to other measures of student achievement in mathematics and English. Furthermore, the researcher and research coordinator was not involved in the administration of the PLE. Therefore, it can only be presumed that the test was administered under the guidelines prescribed by the Uganda National Examinations Board. Future studies should measure student achievements with gain scores across time using several instruments.

Another limitation accrues from the order effect of taking two days’ worth of exams where students are taking the math test first and then the English test on the second day. There may be test fatigue that can impact the accuracy of the results of the English test. Additionally, even though this study used pupil prior achievement as one of the variables, it can be seen that some studies (e.g., Teddlie et al., 2001) show that it is difficult to obtain reliable prior achievement information.
Conclusion

The first objective of this study aimed to create a psychometrically sound measure of teacher quality. A 38-question teacher quality measure was created focusing on four domains: planning and preparation, classroom environment, instruction, and professional responsibilities. The second objective of this study was to examine the relationship between pupil-, teacher-, and school-level factors on pupil achievement on the Primary Leaving Examinations (PLEs) in Uganda. Specifically, sex and SES differences were examined at the pupil and school level, in addition to the influence of teacher quality on pupil exam scores controlling for prior ability.

It was found that that there are significant differences between boys and girls on mathematics achievement controlling for prior ability with boys having higher achievement. Additionally, the results show that there is a significant relationship between student SES (i.e., boarding and day) and student achievement, with boarding school students performing better than their day school counterparts in both subjects. There was also a significant difference between school types (i.e., urban and rural) on student achievement in mathematics, with rural schools performing better than urban schools.

Overall, the purpose of this preliminary study combined across research questions was to begin to validate the scores on a newly developed measure of teacher quality. Ideally, the relationship between sex, pupil SES, school SES, TQM scores, and pupil PLE scores suggests a multilevel influence of the many influences on student achievement. Future research is encouraged to continue to define the theoretical network of
relationships that may exist, and continue to further validate the scores on the measure of teacher quality.
APPENDIX A

ORIGINAL 80-ITEM TEACHER QUALITY MEASURE (TQM)
ORIGINAL 80-ITEM TEACHER QUALITY MEASURE (TQM)

Teacher Quality Measure

<table>
<thead>
<tr>
<th>Rating</th>
<th>Frequency of Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Never (i.e., 0% of the time per week)</td>
</tr>
<tr>
<td>1</td>
<td>Few Occasions (i.e., 1% - 25% of the time per week)</td>
</tr>
<tr>
<td>2</td>
<td>Many Occasions (i.e., 26% - 50% of the time per week)</td>
</tr>
<tr>
<td>3</td>
<td>Most Occasions (i.e., 51% - 75% of the time per week)</td>
</tr>
<tr>
<td>4</td>
<td>Almost All Occasions (i.e., 76% - 100% of the time per week)</td>
</tr>
</tbody>
</table>

Part I: Teacher Lesson Preparation

1. My instructional practices reflect current teaching methods.
2. My instructional goals reflect high expectations related to the curriculum framework.
3. I read about the subject/unit/topic before writing a lesson plan.
4. I demonstrate knowledge of learning resources (e.g., good use of learning resources that make an impression on learning experience, exploring new tools for English teaching such as computers).
5. I acquire knowledge about my pupils’ backgrounds.
6. I contemplate best teaching practices.
7. I develop written lesson plans.
8. My lesson plans do not reflect the syllabi.
9. I arrange pupils’ seats to facilitate monitoring students.
10. My lessons relate to one another to achieve higher student achievements (i.e., developing a link between the previous lesson and the present lesson).
11. My lesson plans have lists of the items needed to aid learning instruction.
12. I use locally available materials for instruction to minimize costs.
13. My lessons are flexible to meet the needs of a variety of pupils in the class (e.g., fast learners, slow learners, etc.).
15. I plan instruction based on curriculum goals.
16. I plan activities that do not promote learning for all pupils.
17. I plan instruction that uses pupils’ strengths as a basis for growth.
Part II: Teacher Classroom Environment

18. My instructional practices reflect current teaching methods.
19. I feel comfortable telling my pupils about my personal experiences.
20. My pupils do not feel comfortable asking for help from me.
21. I have a respectful working relationship with my pupils.
22. I treat each of my pupils fairly.
23. I demonstrate flexibility in responding to pupils (e.g., taking time to listen to each pupil and attend to their unique demands).
24. My pupils have respect for me as a teacher.
25. I maintain a warm learning environment in class (e.g., employing sense of humour, using phrases of encouragement like, “Thank you, John,” or “That was a great answer, Jane.”)
26. When pupils are misbehaving, I give them clear, firm directions to do something (e.g., "Open your book to page 67," or "Please go to your seat now.")
27. I make an effort to promote self-esteem.
28. I give pupils with time to share happenings/events with me.
29. I take the time to share something important with my pupils.
30. I take time to talk about differences/diversity within the classroom (e.g., talking to pupils about respect for other culture).
31. I say no to all forms of bullying.
32. I build activities into the school day that support pupils working together.
33. I focus on the pupils’ strengths when calling upon them to carry out a task.
34. I do not promote respect for each other among my pupils.
35. I take time to educate the class about specific disabilities.
36. I make an effort to promote confidence.
37. I find something positive to say to each pupil.

Part III: Teacher Instruction

38. I begin a lesson with a short review of the previous lesson.
39. I begin a lesson with a short statement of goals.
40. I present new materials in small steps.
41. I allow pupils to practice after each step in the process of learning new material.
42. I provide active practice of lesson material for all pupils.
43. I give short quizzes at the beginning of the lesson.
44. When doing a new lesson, I do not guide pupils during in-class work.
45. I give clear instructions/explanations as I assign in-class work.
46. I circulate around the class during in-class work to monitor performance.
47. I encourage my pupils correcting each other’s in-class work.
48. I have pupils prepare a written summary of the day’s lesson.
49. I do not teach to a variety of learning styles.
50. I use class discussions to help pupils understand the information thoroughly.
51. I provide feedback to pupils in a timely manner.
52. I provide an immediate check for understanding by asking pupils to explain the facts that I have presented.
53. My lesson content is ordered from known to unknown.
54. I use a pupil-centered model of learning.
55. I use formal assessment strategies to evaluate the development of learners (e.g., use of systematic ways to evaluate how well pupils are progressing. For example, after completing a four- to six-week topic, teachers may want to know how well pupils have learned the topic skills and concepts).
56. I use informal assessment strategies to evaluate the development of learners (e.g., assessment may be drawn from typical classroom activities such as assignments, journals, essays, reports, literature discussion groups, or reading logs).
57. I apply the textbook assigned to my English language lesson.
58. I display understanding of subject content matter in my area of specialization (i.e., in-depth knowledge of English).
59. My students understand my verbal lesson goals/strategies.
60. I prepare my pupils for end-of-term exams by assigning class exercises/quizzes.

**Part IV: Teacher Professionalism**

61. I am available to pupils for consultation according to the school policies.
62. I am available to parents for consultation according to the school policies (e.g., during class meetings, visitations days and other days planned by the school).
63. I maintain accurate records of pupils work/progress as well as non-instructional records (e.g., attendance).
64. I demonstrate appropriate conduct while in performance of school responsibilities.
65. I cooperate with colleagues in the school.
66. I do not cooperate with school parents.
67. I am compliant with Uganda’s Teacher Code of Conduct.
68. I adhere to school procedures/rules.
69. I demonstrate ethical behaviour in an educational setting (e.g., teaching, professional responsibilities).
70. I enforce regulations concerning student conduct/discipline.
71. I demonstrate timeliness for assigned responsibilities from administrators.
72. I attend/participate in school activities according to school policies.
73. I assume responsibility for supervising students in out-of-class settings.
74. I cooperate with school administrators.
75. I assume responsibility for the total school program.
76. I demonstrate an attentive attitude during the school day (e.g., classes, meetings).
77. I demonstrate a positive attitude during the school day (e.g., classes, meetings).
78. I demonstrate a respectful demeanour during the school day (e.g., classes, meetings).
79. I conduct classes at their scheduled time.
80. I do not dress appropriately for educational settings (e.g., teaching, professional responsibilities).
APPENDIX B

FINAL 38-ITEM TEACHER QUALITY MEASURE (TQM)
FINAL 38-ITEM TEACHER QUALITY MEASURE (TQM)

Demographic Survey

Directions: This demographic intake contains seven questions about basic demographic characteristics. For each item, please fill in the blank or circle the appropriate response. You can leave any question blank if you do not feel comfortable answering it. Your completion of this survey is your consent to participate.

1. School Name:

2. School Type (Please Circle One):
   
   Rural          Urban

3. Sex (Please Circle One):
   
   Male           Female

4. Age: ______

5. Total Years Teaching (i.e., including the current academic year):
   
   ______________________

6. Current Specialization (e.g., subjects taught):
   
   _____________________________________________________________
   
   ______

7. Highest Certification Earned (i.e., highest level of teacher education; Please Circle One):
   
   Grade III        Grade V        Graduate and Beyond
Staff Survey

**Directions:** This inventory contains 38 items that address specific teacher behaviours. For each item, please circle the appropriate number/rating. The rating scale includes Never, Few Occasions, Many Occasions, Most Occasions, and Almost All Occasions. Your completion of this survey is your consent to participate in this study. If you do not want to participate, simply do not complete the survey. The rating scale is as follows…

<table>
<thead>
<tr>
<th>Rating</th>
<th>Frequency of Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Few Occasions (i.e., 0% - 25% of the time per week)</td>
</tr>
<tr>
<td>1</td>
<td>Many Occasions (i.e., 26% - 50% of the time per week)</td>
</tr>
<tr>
<td>2</td>
<td>Most Occasions (i.e., 51% - 75% of the time per week)</td>
</tr>
<tr>
<td>3</td>
<td>Almost All Occasions (i.e., 76% - 100% of the time per week)</td>
</tr>
</tbody>
</table>

**Part I: Lesson Preparation**

1. My instructional goals reflect high expectations related to the curriculum framework.

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Few</td>
<td>Many</td>
<td>Most</td>
<td>Almost All</td>
</tr>
</tbody>
</table>

2. I demonstrate knowledge of learning resources.

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Few</td>
<td>Many</td>
<td>Most</td>
<td>Almost All</td>
</tr>
</tbody>
</table>

3. I contemplate best teaching practices.

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Few</td>
<td>Many</td>
<td>Most</td>
<td>Almost All</td>
</tr>
</tbody>
</table>

4. My lesson plans do not reflect the syllabi.

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Few</td>
<td>Many</td>
<td>Most</td>
<td>Almost All</td>
</tr>
</tbody>
</table>

5. I arrange pupils’ seats to facilitate monitoring students.

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Few</td>
<td>Many</td>
<td>Most</td>
<td>Almost All</td>
</tr>
</tbody>
</table>
6. My lessons relate to one another to achieve higher student achievements.

0  
Few  

1  
Many  

2  
Most  

3  
Almost All  

7. My lesson plans have lists of the items needed to aid learning instruction.

0  
Few  

1  
Many  

2  
Most  

3  
Almost All  

8. I use locally available materials for instruction to minimize costs.

0  
Few  

1  
Many  

2  
Most  

3  
Almost All  

9. My lessons are flexible to meet the needs of a variety of pupils in the class.

0  
Few  

1  
Many  

2  
Most  

3  
Almost All  

10. I plan instruction based on my knowledge of subject matter.

0  
Few  

1  
Many  

2  
Most  

3  
Almost All  

11. I plan instruction that uses pupils’ strengths as a basis for growth.

0  
Few  

1  
Many  

2  
Most  

3  
Almost All  

Part II: Teacher Classroom Environment

12. My instructional practices reflect current teaching methods.

0  
Few  

1  
Many  

2  
Most  

3  
Almost All  

13. My pupils not feel comfortable asking for help from me.

0  
Few  

1  
Many  

2  
Most  

3  
Almost All
14. I have a respectful working relationship with my pupils.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Few</td>
<td>Many</td>
<td>Most</td>
<td>Almost All</td>
</tr>
</tbody>
</table>

15. I treat each of my pupils equally.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Few</td>
<td>Many</td>
<td>Most</td>
<td>Almost All</td>
</tr>
</tbody>
</table>

16. I take time to talk about differences/diversity within the classroom.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Few</td>
<td>Many</td>
<td>Most</td>
<td>Almost All</td>
</tr>
</tbody>
</table>

17. I focus on the pupils’ strengths when calling upon them to carry out a task.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Few</td>
<td>Many</td>
<td>Most</td>
<td>Almost All</td>
</tr>
</tbody>
</table>

18. I do not promote respect for each other among my pupils.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Few</td>
<td>Many</td>
<td>Most</td>
<td>Almost All</td>
</tr>
</tbody>
</table>

19. I take time to educate the class about specific disabilities.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Few</td>
<td>Many</td>
<td>Most</td>
<td>Almost All</td>
</tr>
</tbody>
</table>

20. I make an effort to promote confidence in my pupils.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Few</td>
<td>Many</td>
<td>Most</td>
<td>Almost All</td>
</tr>
</tbody>
</table>
Part III: Teacher Instruction

21. I present new lesson materials in small steps to allow pupils to practice after each step.

   0          1          2          3
   Few       Many       Most       Almost All

22. I give short quizzes at the beginning of the lesson to check the learning process.

   0          1          2          3
   Few       Many       Most       Almost All

23. I give clear instructions/explanations as I assign in-class work.

   0          1          2          3
   Few       Many       Most       Almost All

24. I encourage my pupils to correct each other’s in-class work.

   0          1          2          3
   Few       Many       Most       Almost All

25. I have pupils prepare a written summary of the day’s lesson.

   0          1          2          3
   Few       Many       Most       Almost All

26. I do not teach to a variety of learning styles.

   0          1          2          3
   Few       Many       Most       Almost All

27. I use class discussions to help pupils understand the information thoroughly.

   0          1          2          3
   Few       Many       Most       Almost All

28. I provide feedback to pupils in a timely manner.

   0          1          2          3
   Few       Many       Most       Almost All
29. My lesson content is ordered from known to unknown.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Few</td>
<td>Many</td>
<td>Most</td>
<td>Almost All</td>
</tr>
</tbody>
</table>

30. I use a pupil-centered model of learning.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Few</td>
<td>Many</td>
<td>Most</td>
<td>Almost All</td>
</tr>
</tbody>
</table>

**Part IV: Teacher Professionalism**

31. I maintain accurate records of pupils’ progress as well as non-instructional records.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Few</td>
<td>Many</td>
<td>Most</td>
<td>Almost All</td>
</tr>
</tbody>
</table>

32. I am compliant with Uganda’s Teacher Code of Conduct.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Few</td>
<td>Many</td>
<td>Most</td>
<td>Almost All</td>
</tr>
</tbody>
</table>

33. I demonstrate ethical behaviour in an educational setting.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Few</td>
<td>Many</td>
<td>Most</td>
<td>Almost All</td>
</tr>
</tbody>
</table>

34. I enforce regulations concerning student discipline.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Few</td>
<td>Many</td>
<td>Most</td>
<td>Almost All</td>
</tr>
</tbody>
</table>

35. I attend/participate in school activities according to school policies.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Few</td>
<td>Many</td>
<td>Most</td>
<td>Almost All</td>
</tr>
</tbody>
</table>

36. I assume responsibility for supervising students in out-of-class settings.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Few</td>
<td>Many</td>
<td>Most</td>
<td>Almost All</td>
</tr>
</tbody>
</table>
37. I demonstrate an attentive attitude during the school day (e.g., classes, meetings).

0 1 2 3
Few Many Most Almost All

38. I do not dress appropriately for educational settings.

0 1 2 3
Few Many Most Almost All
APPENDIX C

APPROVAL TO CONDUCT RESEARCH IN THE WAKISO SCHOOL DISTRICT
The Headteacher,

.................................................................

.................................................................

RE: GATHERING DATA FOR A DOCTORAL PROGRAM

This letter is intended to act as permission for Pius Ochwo to use our District’s student data from the schools’ assessment/National Examination (UPE) for his doctoral program of study (i.e., his dissertation) at Kent State University. He will use the data to discuss the effect of teacher quality on students’ performance. No student or teacher names will be used during this study (i.e., no identifiers).

We look forward to seeing the results of this study.

Lwanga H. Stampijja

For: DISTRICT EDUCATION OFFICER.
APPENDIX D

THE KENT STATE UNIVERSITY INSTITUTIONAL REVIEW BOARD APPROVAL
THE KENT STATE UNIVERSITY INSTITUTIONAL REVIEW BOARD APPROVAL


I am pleased to inform you that the Kent State University Institutional Review Board has reviewed and approved your Application for Approval to Use Human Research Participants as Level I/Exempt research. This application was approved on May 8, 2012. Your research project involves minimal risk to human subjects and meets the criteria for the following category of exemption under federal regulations:

- Exemption 2: Research involving the use of educational tests, surveys, interviews, or observation of public behavior.
- Exemption 4: Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens.

***Submission of annual review reports is not required for Level I/Exempt projects.

If any modifications are made in research design, methodology, or procedures that increase the risks to subjects or includes activities that do not fall within the approved exemption category, those modifications must be submitted to and approved by the IRB before implementation. Please contact the IRB administrator to discuss the changes and whether a new application must be submitted. It is important for you to also keep an unstamped text copy (i.e., Microsoft Word version) of your consent form for subsequent submissions.

Kent State University has a Federal Wide Assurance on file with the Office for Human Research Protections (OHRP); FWA Number 00001853.

If you have any questions or concerns, please contact me by phone at 330-672-2704 or by email at Pwashko@kent.edu.

Respectfully,

Kent State University Office of Research Compliance
224 Cartwright Hall | fax 330.672.2658
Kevin McCrea | Research Compliance Coordinator | 330.672.8058 | kmcrea1@kent.edu
Laurie Kiehl | Research Compliance Assistant | 330.672.0837 | lkiehl@kent.edu
Paulette Washko | Manager, Research Compliance | 330.672.2704 | Pwashko@kent.edu

For links to obtain general information, access forms, and complete required training, visit our website at www.kent.edu/research
RECRUITMENT SCRIPT

Hello. My name is Josephine Nalubowa and I am currently an assistant registrar at Kisubi Brothers University College. My colleague in the United States, Pius Ochwo, and his advisor, are conducting a study investigating the development of a comprehensive, efficient, and portable measure of teacher quality. Completing this survey will assist them in providing invaluable information to other researchers and practitioners.

Participation in this study is completely voluntary. If you decide to participate, your completion of this survey will be your consent.

In this study, you will not have any more risks than you would in a normal day of life. There are no direct benefits by participating in the study. Participation is voluntary, and refusal to take part in the study involves no penalty or loss of benefits to which you are otherwise entitled. You may withdraw from the study at any time. Deciding to participate or not will not impact your standing/relationship to your school or institution. You may leave the study or choose not to complete the survey at any time.

The survey should take no more than 15 to 20 minutes to complete. Please fill out the survey, if you decide to participate, and return the survey to me upon completion. Please do not put your name on the survey, or any other identifying information.

There are a few sections in this survey comprised of demographic and closed-response items. Information obtained from this survey will be analyzed and reported collectively.

For questions, concerns, or complaints about the study, you may contact the Primary Investigator, Pius Ochwo, or the Co-PI, Dr. Aryn C. Karpinski, at pochwo@kent.edu or akarpins@kent.edu or at (330) 672-2012, or the Kent State University IRB, at (330) 672-2704. Thank you for considering to participate in this study!
APPENDIX F

INVITATION TO TAKE PART IN THE STUDY
INVITATION TO TAKE PART IN THE STUDY

You are being invited to participate in a research study. You are being provided with this written statement on the research project, what you will need to do, and the associated risks and benefits of the research. Your participation is voluntary. Please read this carefully. It is important that you ask questions and fully understand the research in order to make an informed decision. This is your copy of the consent information to take with you.

You are invited to participate in a study that will assist researchers in the development of a comprehensive, efficient, and portable measure of teacher quality. The study is being conducted by Pius Ochwo, a doctoral student, and Dr. Aryn C. Karpinski of Kent State University, and it has been approved by the Kent State University Institutional Review Board. No deception is involved, and the study involves no more than minimal risk to participants (i.e., the level of risk encountered in daily life).

Participation in the study typically takes 15 to 20 minutes and is strictly anonymous. You will begin by answering general demographic information about yourself and various questions about teaching behaviors. All responses are treated as confidential, and in no case will responses from individual participants be identified. Rather, all data will be pooled and published in aggregate form only.

In this study, you will not have any more risks than you would in a normal day of life. There are no direct benefits by participating in the study. However, what we learn from the study may help us to better understand teacher quality. Participation is voluntary, refusal to take part in the study involves no penalty or loss of benefits to which participants are otherwise entitled, and participants may withdraw from the study at any time without penalty or loss of benefits to which they are otherwise entitled. Deciding to participate or not will not impact your standing/relationship to your school or institution. You will not be compensated for participating in this study.

If participants have further questions about this study or their rights, or if they wish to lodge a complaint or concern, they may contact Pius Ochwo or Dr. Aryn C. Karpinski, at (330) 672-2012, or pochwo@kent.edu or akarpins@kent.edu. This project has been approved by the Kent State University Institutional Review Board. If you have any questions about your rights as a research participant or complaints about the research, you may call the IRB at (330) 672-2704.

Thank you!
REFERENCES
REFERENCES


Elliot, A. J. (1997). Integrating the “classic” and “contemporary” approaches to achievement motivation: A hierarchical model of approach and avoidance achievement motivation. In M. L. Maehr, & P. R. Pintrich (Eds.), Advances in motivation and achievement (pp.43–179). Greenwich, CT: JAI Press.


www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED497454


Wigfield, A., & Tonks, S. (2002). Adolescents’ expectancies for success and achievement task values during the middle and high school years. In F. Pajares & T. Urdan (Eds.), *Academic Motivation of Adolescents* (pp. 53–82). Greenwich, CT: Information Age.

